Dsouza(10/643,841)
S (mobile or wireless)(3n)(unit?? or device??)
S (transmit? or send?? or transmission??)(3n)signal???)(3n)s1
S (base()station?? or bs or receiver or transceiver??)(3n)s2
S (index??? or value?? or data or information)(3n)(map? or direction??? or location?? or north or south or east or west or N or S or E or W or NE or NW or SE or SW)
S (channel?? or frequency or frequencies)(3n)(matrix??? or cell?? or array??)
S (calculat??? or determine??? or formulat??? or mathematical or determin? or compute or computes or computation)(5n)(s5 or index)
S (power or power()control???? or beam?? or beamforming or beam()(form??? or determination or shaping))(10n)s6
S (chang??? or modify or modification or modifies or improve??? or increas??? or optimi? or strength? or redirect???)(3n)s7
s au=(Lau, K? or Lau K?)

RESULTS OF THE TSUNAMI FIELD TRIALS:
POSITION LOCATION IN MACRO AND MICRO CELL ENVIRONMENTS

P T Thompson, D Brooks[1]

Introduction

The CEC ACTS TSUNAMI (II) project involved the development, integration and field trial evaluation of a DCS1800 2nd generation mobile base-station with an adaptive antenna. The consortium[2] designed and built a 8 element transmit and receiving adaptive antenna system, employing digital beamforming techniques operating over a DCS1800 air interface and the performance assessed for a variety of different operational scenarios using the Orange testbed facility in Bristol. The TSUNAMI (II) field trial system included both transmit and receive calibration facilities in order to obtain adequate accuracy from the digital beamforming techniques. The ability to track users is a fundamental feature of this technology.

The application of array signal processing to provide mobile location information has been assessed, using the data collected during the trials and compared with position information, obtained via a commercially available GPS receiver. This application is particularly relevant to 'hot-spot' detection in cellular networks as well as the '911' location requirement in the US.

Mobile location

The problem of accurate position estimation of cellular subscribers is receiving growing attention for a number of reasons. In the United States regulations are to be introduced which will require wireless operators to locate accurately subscribers making emergency calls ('911'). Network operators could also use subscriber location information for cellular planning (i.e. traffic 'hot spot' detection), location sensitive billing and also for offering new services. This type of facility can be provided through the use of array signal processing and can be regarded as a *value added service* if the array is also providing capacity enhancements to the network.

A number of technical solutions have been proposed to this problem. Most of these techniques rely on measuring the time of arrival (TOA) of the signal from the mobile station at three or more receiving points, from which the location of the mobile station can be calculated. Other techniques use direction of arrival (DOA) measurements from at least two receiving locations. Here an alternative technique using an antenna array at a single receiving point to estimate mobile transmitter location by combining time of arrival and DOA information is proposed. This technique has the advantage that, in some applications (particularly traffic hot spot detection), only one active receiving site is required, and furthermore this can operate independently to the rest of the communications network.

This paper reports on work carried out by ERA Technology on results obtained by using mobile location algorithms on the TSUNAMI (II) field trial data which was conducted as an adjunct to the TSUNAMII work packages.

---

[1] The authors are with ERA Technology Ltd.
[2] Consortia comprising: ERA Technology (coordinator), Motorola ECID, Orange PSC, University of Bristol, Wireless Systems International, CASA, Robert Bosch, University of Aalborg, France Telecom CNET, University Polytechnic of Catalunya.

Algorithm Description

In the proposed algorithm, the processing for source location consists of three key stages. Initially, the channel impulse response for each element of the antenna array is estimated from the signal vector received at the base station. Then, estimates of the direction of arrival (DOA) and ranges are extracted. Finally, these raw estimates are filtered to obtain the estimates of mobile trajectory. This approach has a number of advantages, in particular:

- The method is simple and fast, since it is based on channel impulse response estimates and is therefore suitable for real-time applications.

- The influence of interference from other cells is limited, since channel impulse responses are discriminated by means of the training sequence de-correlation properties.

- Post-processing is applied to estimates to increase position accuracy.

Each GSM burst contains a 26-bit training sequence, shared by all mobiles of a given cell. Interference from other cells is limited by using orthogonal training sequences. When modulated, the training sequence produces a 16-bit sequence, periodically extended by 5 bits to either side. When MSK modulated, the 16-bit sequence has an ideal (Dirac) auto-correlation function. GMSK modulation broadens the main lobe and introduces side lobes.

Channel impulse responses, H, are estimated and stored at the basestation equipped with the appropriate array and signal processing hardware. This is simply achieved by cross-correlating the modulated training sequence, corresponding to the relevant mobile station, with the received burst. When channel distortions have been introduced by multipath, H is a weighted sum of Dirac functions. In order to improve the time resolution of the paths, the impulse responses can be interpolated. Finally, H is pre-multiplied by the beamformer weight coefficients matrix, W, in order to obtain the spatial impulse response for the antenna array, H_θ. The squared magnitudes of the entries in H_θ represent the received energy from the burst, as a function of time of arrival and direction of arrival. The dominant multipath components are estimated from the highest values of H_θ. DOA estimates are given by the row index of the peaks, and time of arrival estimates by the corresponding column index. The range estimates are found by multiplying the time of arrival estimates by the speed of light. In the following analysis, only one peak is selected from each spatial impulse response estimate obtained from the measured data.

The discrete nature of the matrix W imposes the DOA accuracy of one degree steps, covering –60° to +60° with respect to the array boresight. Range accuracy is limited by the number of samples per bit used in the interpolated channel impulse response estimation, and here 32 samples per bit have been used. The range resolution is approximately 17m, since the sidelobes of the autocorrelation function of the training sequence tend to bias these estimates.

Post-processing was employed to reduce quantization noise in the location estimates. Here, DOA and range estimates are made at regular intervals. Then, histograms of the DOA and range estimates are computed. These histograms are almost always mono-modal, except in the presence of interference, which introduces a smaller second peak in the histogram. The values corresponding to the histogram maxima only are used as representative estimates for the interval if their relative frequencies are greater than a predefined threshold. Otherwise, representative estimates are obtained from the linear extrapolations of the previous values.

Macrocell Results

The algorithms have been tested for one stationary test, with line-of-sight to the antenna mast, and seventeen moving tests.

Results are reported in Table 1, where angles are given in degrees, distances in metres and error probabilities in percentage. Figure 1 represents the true and estimated trajectories for the test 045.

Test	DOA Error degs		Range Error m		Position Error, m	P(err<Xm) %		
	Mean	Std	Mean	Std	RMSE	X=50m	X=125m	X=250m
000	0.0	0.5	25	74	88	47.0	87.1	100.0
010	0.5	2.4	39	154	236	27.8	73.0	92.2
019	-1.0	2.0	57	143	387	27.8	52.4	67.8
024	-14.0	34.0	-36	154	1917	15.4	37.6	70.6
033	-12.8	34.3	22	95	2003	13.3	46.1	85.2
039	0.0	0.6	-4	104	118	15.8	71.1	100.0
041	0.0	0.4	0	61	67	64.8	100.0	100.0
045	-0.1	0.5	12	91	111	41.2	82.9	100.0
047	-0.5	1.6	-51	167	266	27.8	57.0	81.6
070	0.1	0.9	5	122	143	27.7	64.2	96.1
089	0.0	0.6	389	978	1055	15.6	46.5	74.5
108	-7.6	21.0	12	110	1288	11.1	60.5	84.0
115	-0.1	1.8	-2	112	165	23.9	55.5	93.3
119	0.0	0.4	-25	97	118	25.1	73.1	98.8
125	-0.8	2.2	-10	141	203	11.9	49.7	82.0
129	-24.8	38.5	95	163	2704	14.2	45.2	64.2
139	-0.5	1.6	-22	140	182	27.1	67.0	82.4
142	-8.7	23.1	-8	107	1415	24.6	55.9	81.6

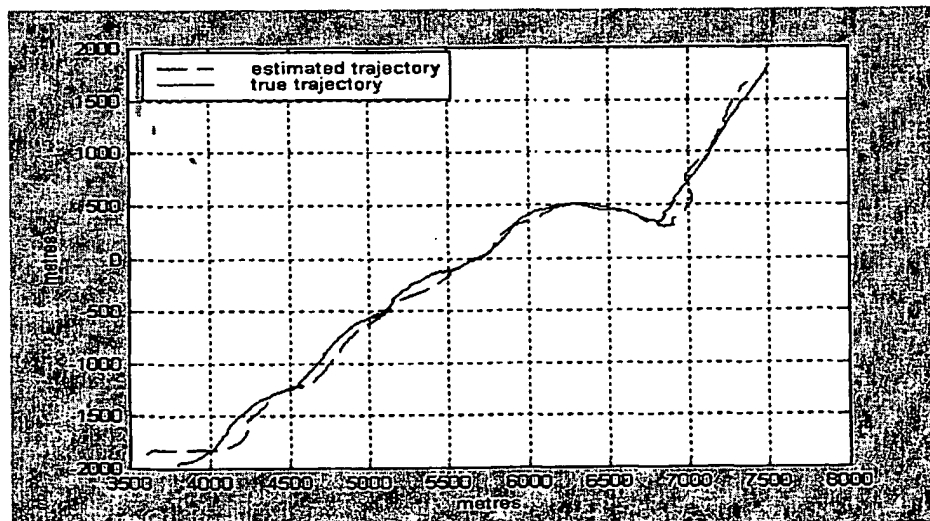**Table 1: Results for the macrocell trials**

Figure 1 : GPS (solid line) and estimated (dashed line) positions for the test 045. The base station is at the origin of the figure.

Macro-cell location performance

In Table 1, location performance has been evaluated in terms of the mean and standard deviations of the DOA, range errors, as well as the root mean square position error (RMSE), for both estimated and GPS derived position location. Furthermore, the accuracy of the location estimates is assessed by counting the number of position estimates within increasing radii circles (50m, 125m, and 250m) centred on the true positions. Since differential GPS equipment was not available during the tests, the GPS measurements are subject to a bias that varies between experiments. When the exact routes are known, the measured GPS positions are overlaid onto an Ordnance Survey map and the GPS trajectories are shifted to be as close as possible to the exact routes. Thus, the results include the influence of any short-term GPS positioning errors.

Macro-cell position accuracy

The following parameters were found to be particularly sensitive in terms of the resultant position accuracy:

- The number of samples per bit and the number of angles used in the channel impulse response estimation, gives an uncertainty on estimates of 35m and 1°. Filtering was introduced to decrease this effect.
- Drift in frequency in the mobile station, this puts a lower limit on the range variance of approximately 137 metres.
- Drift in frequency has no impact on the DOA estimates, but angle errors dominate in the macro tests: A variation of 1° at 5km corresponds to 75m.
- Errors in location system calibration add bias terms that are difficult to estimate.

Some problems have been identified for some of the test runs.

- At the end of two test runs the interference was selected instead of the signal.

- For three tests the estimated trajectory moves away from the true position, probably because of the extrapolation method used between two *reliable* points.

- At the ends of some test runs the angles are greater than +60°. The mobiles are then out of range and the DOA estimates correspond to side lobes of the channel impulse responses.

For the remaining nine tests, the average RMS error is 136m. On average, the estimated positions are within 50m of the true positions one third of the time, within 125m three quarters of the time and 96% of the estimates are within 250m of the true positions.
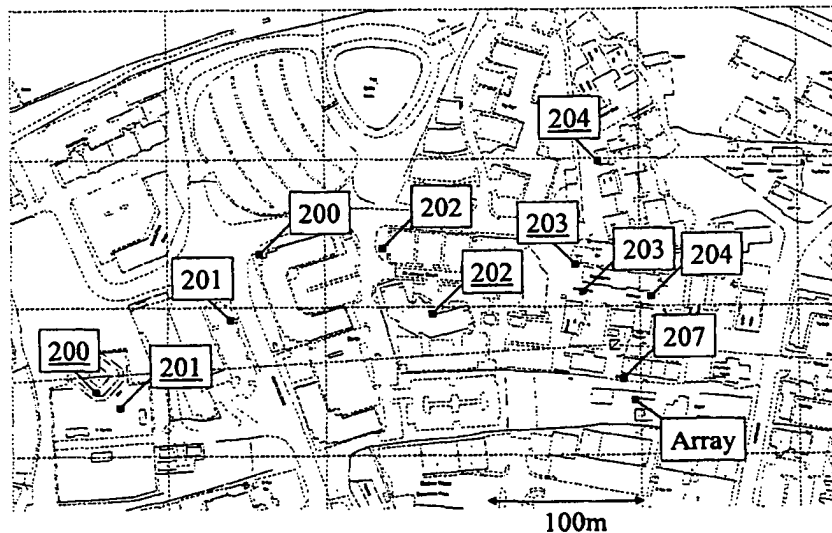
Microcell Results

Here the algorithm was applied to six stationary tests as well as five moving ones.

For the static tests the general results are given in Table 2 where angles are given in degrees, distances in metres and error probabilities in percentage. Figure 2 represents the exact positions and the mean positions of the estimates on the map.

Test	DOA Error degs		Range Error m		Position Error, m	Spread m		P(err<Xm) %		
	Mean	Std	Mean	Std	RMSE	X	Y	X=50	X=125	X=250
200	21.4	0.5	-463	35	456	85	100	5.9	5.9	5.9
201	12.6	0.3	69	31	99	95	60	11.8	100.0	100.0
202	7.5	0.0	-44	35	60	102	50	68.6	100.0	100.0
203	-4.0	0.1	25	43	50	137	46	78.4	100.0	100.0
204	17.9	5.6	102	20	110	327	584	3.9	100.0	100.0
207	0.0	0.4	0	41	41	129	65	84.8	100.0	100.0

Table 2: Results for the microcell, stationary tests.

- For test 204, range estimates are always around the same average value, but the DOA estimates take values around −21 degrees or −45 degrees, probably corresponding to two different reflections.
- The frequency/time drift inside the mobile terminal is particularly visible and puts a limit on the range accuracy. This time drift implies shifts of one fourth of a bit period for the travel from the base station to the mobile and back to the base station, i.e. a corresponding range shift of 137m approximately.
- Values for the standard deviations are generally similar to those obtained for the macro tests.
- Mean values are larger than those obtained with the macro tests. These large biases are mainly due to reflections. Exploiting geometric properties of reflections may reduce them.
- Reflections modify angles and ranges.
- Exact positions from reliable estimates or positions of the mobile as seen by the basestation can be determined from the building positions and heights.
- However, taking into account reflections increases the region of uncertainty of estimates. Using several reflections for the same test may reduce this uncertainty.


Figure 2: True and estimated (underlined) positions for the stationary microcell tests

The five moving tests all followed the same route (see Figure 4). Here, the results are very similar, showing that the experiment is repeatable. After 400 seconds, the mobile is out of range since its angle of arrival is less than –60 degrees. Only the first parts of the tests are thus considered. Figure 3 represents the DOA and range estimates for the test 277. In the histogram computation, all representative estimates were kept and the averaging filter has not been applied.

DOA estimates present clear changes of slopes. They may correspond to different reflections and could be solved with a map, indicating building positions and heights. Range estimate slopes also present discontinuities that are not always easily associated with the reflections defined from the angle estimates and may be involved by the time drift inside the mobile terminal.

The lines in Figure 3 indicate the changes in the slopes of the DOA for the test 277. The same lines are drawn for the range estimates. The positions on the trajectory at the corresponding times are represented by large dots in Figure 4. Many of them are explained by the building locations, even without knowing their heights. With a three-dimensional map, more precise estimates of the mobile location may be achievable.

Figure 5 indicates the position estimates derived for a few fixed locations in more detail. The influence of multipath propagation from surrounding buildings is clearly evident.
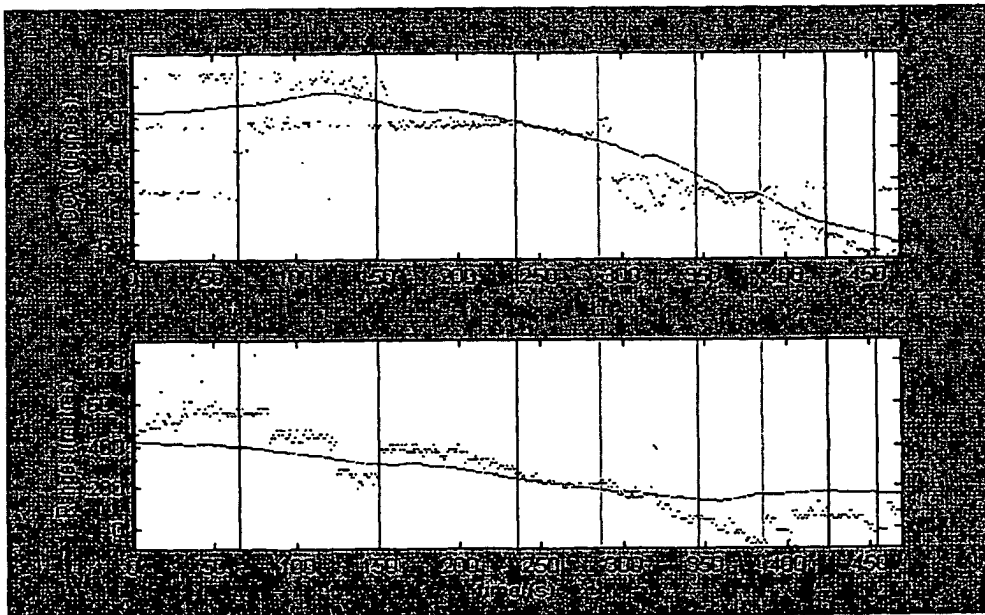
Figure 3: True (solid line) and estimated (dotted line) DOA and ranges for the test 277
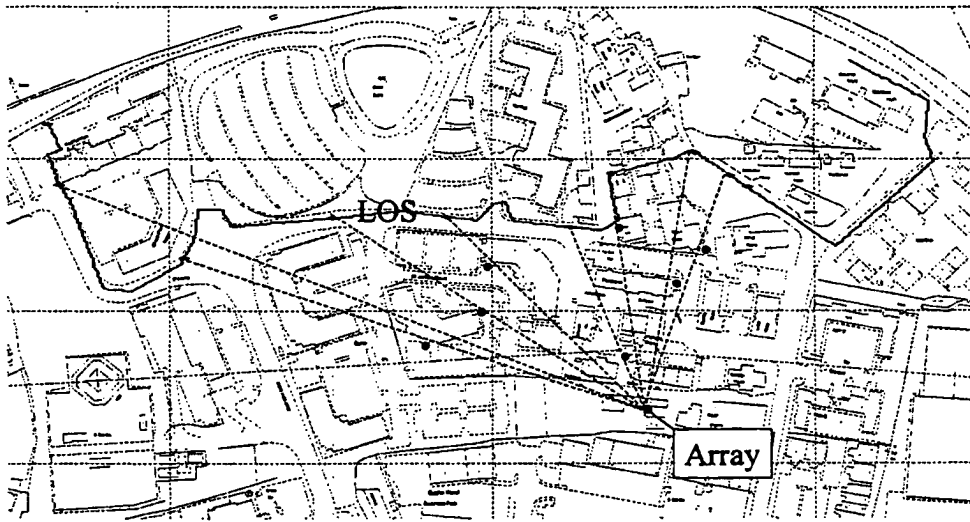
Figure 4: Route for the moving micro tests.
Points on the trajectory correspond to the lines in Figure 3

Conclusions

A technique for estimating the location of a mobile radio transmitter, using an antenna array at a single receiving point has been presented. Such a technique could be used to aid cell planning in a cellular network by mapping traffic density within a coverage area, or as a solution to the US 911 location requirement.

The technique has been tested using the TSUNAMI (II) macrocell and microcell field trial data. The macrocell results show that the mobile can be located to within a circle of radius 125m for approximately 75% of the time. It is believed that a large component of the position error is due to the timing uncertainty within the mobile itself, which can be up to one quarter of a bit period.

The microcell environment is clearly more challenging, because the perceived direction of arrival is often significantly different from the true bearing of the mobile station, resulting in biased estimates of DOA and range. The overall position accuracy in the microcell tests appeared to be better than the macrocell case. However, this is not a fair comparison since the position accuracy for the microcell was only evaluated for a limited number of stationary tests. In the moving microcell tests, sudden changes in the channel (the so called 'corner effect') can be clearly seen in the DOA and range estimates and in some cases these changes can be clearly related to the positions of buildings within the test environment. Nonetheless, the microcell results demonstrate that the technique has promise but requires further refinement to ameliorate the influence of multipaths.
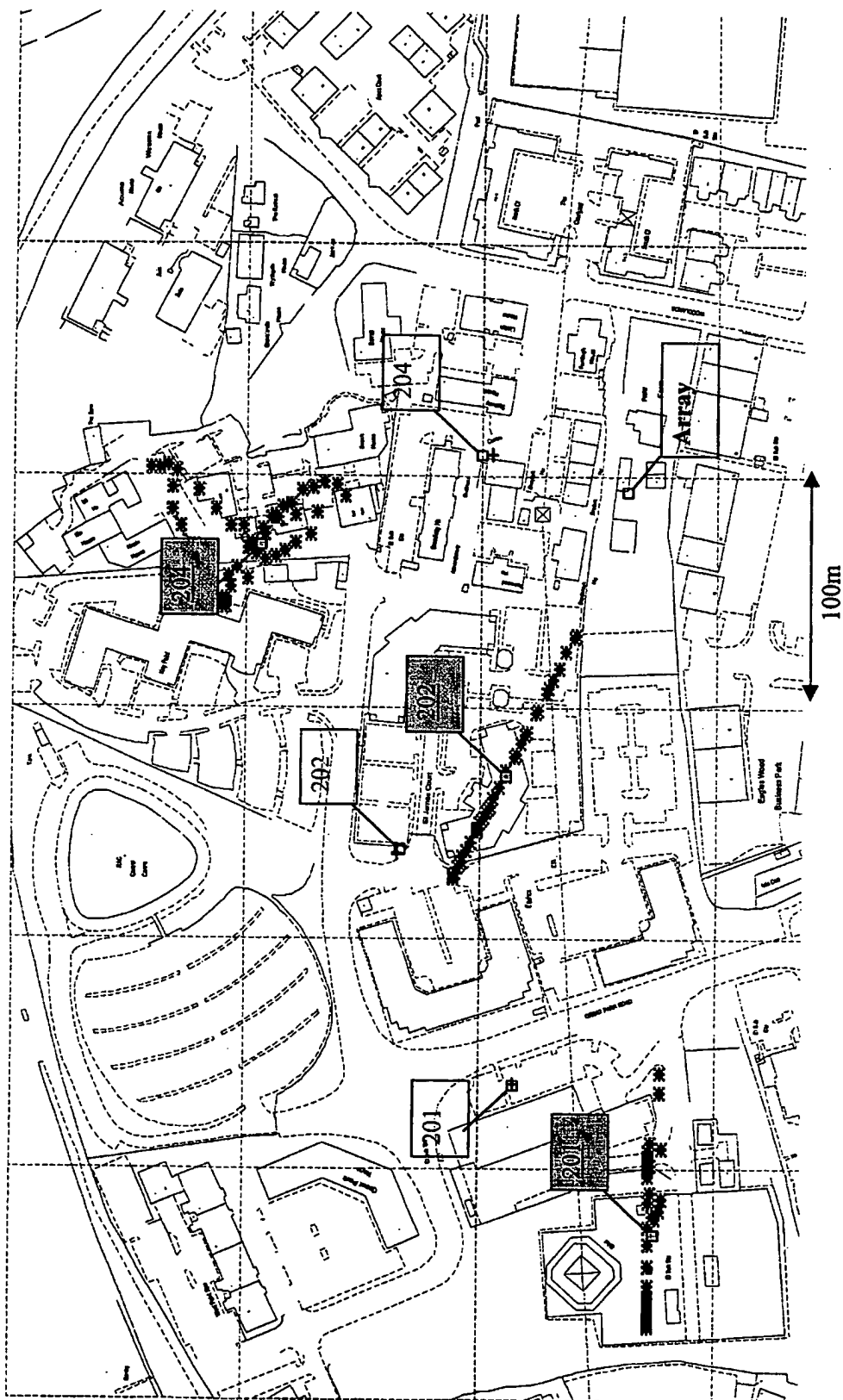
3/7

Figure 5 Plot of the actual position and the estimated (underlined) position for a few cases in the Micro-cell trials.

100m

Data Gathering Algorithms in Sensor Networks Using Energy Metrics

Stephanie Lindsey, Cauligi Raghavendra, *Fellow Member, IEEE,* and
Krishna M. Sivalingam, *Senior Member, IEEE*

Abstract—Sensor webs consisting of nodes with limited battery power and wireless communications are deployed to collect useful information from the field. Gathering sensed information in an energy efficient manner is critical to operating the sensor network for a long period of time. In [12], a data collection problem is defined where, in a round of communication, each sensor node has a packet to be sent to the distant base station. There is some fixed amount of energy cost in the electronics when transmitting or receiving a packet and a variable cost when transmitting a packet which depends on the distance of transmission. If each node transmits its sensed data directly to the base station, then it will deplete its power quickly. The LEACH protocol presented in [12] is an elegant solution where clusters are formed to fuse data before transmitting to the base station. By randomizing the cluster-heads chosen to transmit to the base station, LEACH achieves a factor of 8 improvement compared to direct transmissions, as measured in terms of when nodes die. An improved version of LEACH, called LEACH-C, is presented in [14], where the central base station performs the clustering to improve energy efficiency. In this paper, we present an improved scheme, called PEGASIS (Power-Efficient GAthering in Sensor Information Systems), which is a near-optimal chain-based protocol that minimizes energy. In PEGASIS, each node communicates only with a close neighbor and takes turns transmitting to the base station, thus reducing the amount of energy spent per round. Simulation results show that PEGASIS performs better than LEACH by about 100 to 200 percent when 1 percent, 25 percent, 50 percent, and 100 percent of nodes die for different network sizes and topologies. For many applications, in addition to minimizing energy, it is also important to consider the delay incurred in gathering sensed data. We capture this with the $energy \times delay$ metric and present schemes that attempt to balance the energy and delay cost for data gathering from sensor networks. Since most of the delay factor is in the transmission time, we measure delay in terms of number of transmissions to accomplish a round of data gathering. Therefore, delay can be reduced by allowing simultaneous transmissions when possible in the network. With CDMA capable sensor nodes [11], simultaneous data transmissions are possible with little interference. In this paper, we present two new schemes to minimize $energy \times delay$ using CDMA and non-CDMA sensor nodes. If the goal is to minimize only the delay cost, then a binary combining scheme can be used to accomplish this task in about $\log N$ units of delay with parallel communications and incurring a slight increase in energy cost. With CDMA capable sensor nodes, a chain-based binary scheme performs best in terms of $energy \times delay$. If the sensor nodes are not CDMA capable, then parallel communications are possible only among spatially separated nodes and a chain-based 3-level hierarchy scheme performs well. We compared the performance of direct, LEACH, and our schemes with respect to $energy \times delay$ using extensive simulations for different network sizes. Results show that our schemes perform 80 or more times better than the direct scheme and also outperform the LEACH protocol.

Index Terms—Wireless sensor networks, data gathering protocols, energy-efficient operation, greedy algorithms, performance evaluation.

◆

1 INTRODUCTION

INEXPENSIVE sensors capable of significant computation and wireless communications are becoming available [4], [6], [8], [10], [16], [23]. A web of sensor nodes can be deployed to collect useful information from the field in a variety of scenarios including military surveillance, landmine detection, in harsh physical environments, for scientific investigations on other planets, etc. [1], [10], [16], [29]. These sensor nodes can self-organize to form a network and can communicate with each other using their wireless interfaces. Energy efficient self-organization and initialization protocols are developed in [18], [19]. Each node has transmit power control and an omni-directional antenna, and therefore can adjust the area of coverage with its wireless transmission. Typically, sensor nodes collect audio, seismic, and other types of data and collaborate to perform a high-level task in a sensor web. For example, a sensor network can be used for detecting the presence of potential threats in a military conflict. Since wireless communications consume significant amounts of battery power, sensor nodes should be energy efficient in transmitting data [3], [17], [25], [27]. Energy efficient communication in wireless networks is attracting increasing attention in the literature [5], [22], [24], [28], [30].

A typical application in a sensor web is gathering of sensed data at a distant base station (BS) [12]. Fig. 1 shows a 100-node sensor network in a playing field of size $50m \times 50m$. There is an energy cost for transmitting or receiving a packet in the radio electronics and there is a variable energy cost depending on the distance in transmissions. Due to the r^2 .or larger radio signal attenuation for a

- S. Lindsey is with Microsoft Corporation, Redmond, WA 98052.
- C. Raghavendra is with the Department of Electrical Engineering Systems, University of Southern California, Los Angeles, CA 90089-2562, and with the Computer Systems Research Department, The Aerospace Corporation, PO Box 92957, Los Angeles, CA 90009-2957, E-mail: raghu@usc.edu
- K.M. Sivalingam is with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164-2752.
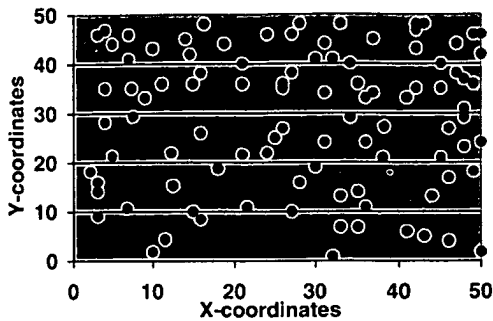
Fig. 1. Random 100-node topology for a $50m \times 50m$ network. The base station (BS) is assumed to be located at (25, 150), which is at least 100m from the nearest node.

range r, it is important to limit transmission distances to conserve energy.

In this paper, we assume the following:

- Each sensor node has power control and the ability to transmit data to any other sensor node or directly to the BS [20], [22].
- Our model sensor network contains homogeneous and energy constrained sensor nodes with initial uniform energy.
- Every node has location information.
- There is no mobility.

1.1 Energy Reduction for Data Gathering in Sensor Networks

In each round of this data-gathering application, all data from all nodes need to be collected and transmitted to the BS, where the end-user can access the data. In some sensor network applications, data collection may be needed only from a region and, therefore, a subset of nodes will be used. A simple approach to accomplishing this data gathering task is for each node to transmit its data directly to the BS. Since the BS is typically located far away, the cost to transmit to the BS from any node is high so nodes will die very quickly. Therefore, an improved approach is to use as few transmissions as possible to the BS and reduce the amount of data that must be transmitted to the BS in order to reduce energy. Further, if all nodes in the network deplete their energy levels uniformly, then the network can operate without losing any nodes for a long time.

In sensor networks, data fusion helps to reduce the amount of data transmitted between sensor nodes and the BS [9], [15], [31]. Data fusion combines one or more data packets from different sensor measurements to produce a single packet, as described in [12]. For example, sensors may collect temperature, pressure, humidity, and signal data from the field. We would be interested in finding the maximum or minimum values of such parameters. Data fusion can be used here to combine one or more packets to produce a same-size resultant packet. The LEACH protocol presented in [12] is an elegant solution to this data collection problem where a small number of clusters are formed in a self-organized manner. The nice property of the LEACH protocol is that it is completely distributed and sensor nodes organize in a cluster hierarchy to fuse their data to eventually transfer to the BS. In LEACH, a designated node in each cluster collects and fuses data from nodes in its cluster and transmits the result to the BS. LEACH uses randomization to rotate the cluster heads and

achieves a factor of eight improvement compared to the direct approach, before the first node dies.

In LEACH, clusters are formed in a self-organized manner in each round of data collection. About 5 percent of the nodes in the network selected randomly become cluster heads. These cluster heads send a strong beacon signal to all nodes and sensor nodes decide which cluster to join based on received signal strength. The distributed cluster formation in each round in LEACH may not produce good clusters to be efficient. In an improved version of this scheme, called LEACH-C [14], this cluster formation is done at the beginning of each round using a centralized algorithm by the BS. Although the energy cost for cluster formation is higher in LEACH-C, the overall performance is better than LEACH due to improved cluster formation by the BS. The steady state part of the LEACH-C protocol, i.e., data collection in rounds, is identical to the LEACH protocol (p. 94 in [14]). LEACH-C improves the performance by 20 percent to 40 percent (p. 97 in [14]), depending on the network parameters, compared to LEACH in terms of the total number of rounds of data collection that can be achieved before sensor nodes start to die.

Further improvements can be obtained if each node communicates only with close neighbors and only one designated node sends the combined data to the BS in each round in order to reduce energy. A new protocol based on this approach, called PEGASIS (Power-Efficient GAthering in Sensor Information Systems), is presented in this paper, which significantly reduces energy cost to increase the life of the sensor network. The PEGASIS protocol is near optimal in terms of energy cost for this data gathering application in sensor networks. The key idea in PEGASIS is to form a chain among the sensor nodes so that each node will receive from and transmit to a close neighbor. Gathered data move from node to node, get fused, and, eventually, a designated node transmits to the BS. Nodes take turns transmitting to the BS so that the average energy spent by each node per round is reduced. Building a chain to minimize the total length is similar to the traveling salesman problem, which is known to be intractable. However, with the radio communication energy parameters, a simple chain built with a greedy approach performs quite well. The PEGASIS protocol achieves between 100 to 200 percent improvement when 1 percent, 25 percent, 50 percent, and 100 percent of nodes die compared to the LEACH protocol. PEGASIS performance improvement in comparison with LEACH-C will be slightly less as LEACH-C improves upon LEACH by about 20 percent to 40 percent. In the rest of this paper we present all our performance comparisons with respect to the LEACH protocol with the understanding that the improvement is less by the extent that LEACH-C improves upon LEACH [14]. When attribute-based search is to be performed, then the area and, hence, selected sensor nodes, will also change dynamically. In these situations, the BS selects the area of interest and only selected nodes in the region participate in data collection. We will still use the same chain ordering of nodes and only the selected nodes will be on to form the truncated chain. Likely, these nodes will still be nearby on the shortened chain and the data collection will still be efficient.

Our scheme can be modified appropriately if some of the stated assumptions about sensor nodes are not valid. If nodes are not within transmission range of each other, then alternative, possibly multihop transmission paths will have

to be used. In fact, our chain-based schemes will not be affected that much as each node communicates only with a local neighbor and we can use a multihop path to transmit to the BS. We need to make some adjustments in the chain construction procedure to ensure that no node is left out. Other schemes, including LEACH, rely on direct reachability to function correctly. To ensure balanced energy dissipation in the network, an additional parameter could be considered to compensate for nodes that must do more work every round. If the sensor nodes have different initial energy levels, then we could consider the remaining energy level for each node in addition to the energy cost of the transmissions. The assumption of location information is not critical. The BS can determine the locations and transmit to all nodes or the nodes can determine this through received signal strengths. For example, nodes could transmit progressively reduced signal strengths to find a close neighbor to exchange data. This would require the nodes to consume some energy when trying to find local neighbors; however, this is only a fixed initial energy cost when constructing the chain. If nodes are mobile, then different methods of transmission could be examined. For instance, if nodes could approximate how often and at what speed other nodes are moving, then it could determine more intelligently how much power is needed to reach the other nodes. Perhaps, the BS can help coordinate the activities of nodes in data transmissions. Discussion of schemes with mobile sensor nodes is beyond the scope of this paper.

1.2 *Energy × Delay* Reduction for Data Gathering in Sensor Networks

Another important factor to consider in the data gathering application is the average delay per round. Here, we assume that data gathering rounds are far apart and the only traffic in the network is due to sensor data. Therefore, data transmissions in each round can be completely scheduled to avoid delays in channel access and collisions. The delay for a packet transmission is dominated by the transmission time as there is no queuing delay and the processing and propagation delays are negligible compared to the transmission time. With the direct transmission scheme, nodes will have to transmit to the base station one at a time, making the delay a total of N units (one unit per transmission, where N is equal to the number of nodes). To reduce delay, one needs to perform simultaneous transmissions. The well-known approach of using a binary scheme to combine data from N nodes in parallel will take about $\log N$ units of delay, although incurring an increased energy cost. *Energy × delay* is an interesting metric to optimize per round of data gathering in sensor networks.

Why *energy × delay* metric? Clearly, minimizing energy or delay in isolation has drawbacks. For battery operated sensors, longevity is a major concern and priorities can be entirely different when energy reserves become depleted. Energy efficiency often brings additional latency along with it. Minimizing delay is not always practical in sensor network applications. Maximizing the throughput is not the best strategy for energy-critical links. Generally, increased energy savings come with a penalty of increased delay. However, several practical applications set limits on acceptable latency, as specified by QoS requirements. For example, the data gathering delay per round may have a bound. Therefore, there is a tradeoff between energy spent per packet and delay; *energy × delay* is an appropriate

measure to optimize for in wireless sensor networks. Specifically, our view is that minimizing *energy × delay* while meeting acceptable delays for applications can lead to significant power savings.

Simultaneous wireless communications among pairs of nodes is possible only if there is minimal interference among different transmissions. CDMA technology can be used to achieve multiple simultaneous wireless transmissions with low interference. If the sensor nodes are CDMA capable, then it is possible to use the binary scheme and perform parallel communications to reduce the overall delay. However, the energy cost may have to go up slightly as there will still be a small amount of interference from other unintended transmissions. Alternatively, with a single radio channel and non-CDMA nodes, simultaneous transmissions are possible only among spatially separated nodes. Since the energy costs and delay per transmission for these two types of nodes are quite different, we will consider *energy × delay* reduction for our data gathering problem separately for these two cases.

In this paper, we present the following new protocols for data gathering using the *energy × delay* metric:

- a binary chain-based scheme with CDMA sensor nodes,
- a three level chain-based scheme which performs better than direct and PEGASIS with this metric for non-CDMA sensor nodes.

Both of these protocols use hierarchical organization of sensor nodes with possible simultaneous data transmissions in each level of the hierarchy. A greedy chain is formed among the sensor nodes in both of these protocols which will form the lowest level in the hierarchy. The binary scheme has a hierarchy of $\lceil \log N \rceil$, where N is the number of nodes in the sensor network. The second protocol uses a 3-level hierarchy by forming groups in each level and promoting one node from each group to the next level. Simulation results show that both schemes perform 80 or more times better than direct scheme and the binary scheme performs eight times better than LEACH with respect to the *energy × delay* metric.

This paper is organized as follows: In Section 2, the radio model for energy calculations used throughout this paper is discussed. In Section 3, an analysis of the energy cost is given for the data gathering problem. The PEGASIS scheme is presented in Section 4, which is shown to be a near-optimal solution for minimizing energy. In Section 5, an analysis of the *energy × delay* metric for data gathering is given. Two new protocols for reducing *energy × delay* for data gathering with and without CDMA capable nodes are presented in Sections 6 and 7, respectively. Extensive simulation results with different size networks and simulation parameters are presented in Section 8. In all our simulation experiments, we considered only the original LEACH protocol and our proposed new protocols. The performance improvements with respect to LEACH-C will be slightly less corresponding to the extent LEACH-C improves upon LEACH. Finally, some concluding remarks are given in Section 9.

2 RADIO MODEL FOR ENERGY CALCULATIONS

We use the same radio model as discussed in [12], which is the first order radio model. In this model, a radio dissipates $E_{elec} = 50nJ/bit$ to run the transmitter or receiver circuitry

and $\epsilon_{amp} = 100pJ/bit/m^2$ for the transmitter amplifier. The radios have power control and can expend the minimum required energy to reach the intended recipients. The radios can be turned off to avoid receiving unintended transmissions. An r^2 energy loss is used due to channel transmission [21], [26]. The equations used to calculate transmission costs and receiving costs for a k-bit message and a distance d are shown below:

.0.a Transmitting

$$E_{Tx}(k, d) = E_{Tx-elec}(k) + E_{Tx-amp}(k, d)$$

$$E_{Tx}(k, d) = E_{elec} \times k + \epsilon_{amp} \times k \times d^2$$

.0.b Receiving

$$E_{Rx}(k) = E_{Rx-elec}(k)$$

$$E_{Rx}(k) = E_{elec} \times k$$

Receiving data is also a high cost operation, therefore, the number of receptions and transmissions should be minimal to reduce the energy cost of an application. With these radio parameters, when $k = 2,000$ and d^2 is 500, the energy spent in the amplifier part equals the energy spent in the electronics part and, therefore, the cost to transmit a packet will be twice the cost to receive. It is assumed that the radio channel is symmetric so that the energy required to transmit a message from node i to node j is the same as the energy required to transmit a message from node j to node i for a given signal-to-noise ratio (SNR), typically 10 dB. For the comparative evaluation purposes of this paper, we assume that there are no packet losses in the network. It is not difficult to model errors and losses in terms of increased energy cost per transmissions. With known channel error characteristics and error coding, this cost can be modeled by suitably adjusting the constants in the above equations.

When there are multiple simultaneous transmissions, the transmitted energy should be increased to ensure that the same SNR as with a single transmission is maintained. With CDMA nodes using 64 or 128 chips per bit (which is typical), the interference from other transmissions is calculated as a small fraction of the energy from other unintended transmissions. This effectively increases the energy cost to maintain the same SNR. With non-CDMA nodes, the interference will equal the amount of energy seen at the receiver from all other unintended transmitters. Therefore, only a few spatially distant pairs can communicate simultaneously in the network.

3 ENERGY COST ANALYSIS FOR DATA GATHERING

In this section, we will analyze the energy cost of data gathering from a sensor web to the distant BS. Recall that the data collection problem of interest is to gather a k-bit packet from each sensor node in each round. Of course, the goal is to keep the sensor web operating as long as possible. A fixed amount of energy is spent in receiving and transmitting a packet in the electronics and an additional amount proportional to d^2 is spent while transmitting a packet. There is also a cost of 5 nJ/bit/message for 2,000 bit messages in data fusion. With the direct approach, all nodes transmit directly to the BS, which is usually located at some

distance from the sensor network. Therefore, every node will consume a significant amount of power to transmit to the BS in each round. Since the nodes have a limited amount of energy, nodes will die quickly, causing the reduction of the system lifetime.

As observed in [12], the direct approach would work best if the BS is located close to the sensor nodes or the cost of receiving is very high compared to the cost of transmitting data. For the rest of the analysis, we use 50, 100, and 200-node sensor networks. In a scenario where the BS is located far away, energy costs can be reduced if the data is gathered locally among the sensor nodes and only a few nodes transmit the fused data to the BS. This is the approach taken in LEACH and its variants, where clusters are formed dynamically in each round and cluster-heads (leaders for each cluster) gather data locally and then transmit to the BS. Cluster-heads are chosen randomly, but all nodes have a chance to become a cluster-head in LEACH to balance the energy spent per round by each sensor node. For a 100-node network in a $50m \times 50m$ field with the BS located at (25, 150), which is at least 100 meters from the closest node, LEACH achieves a factor of 8 improvement compared to the direct approach in terms of number of rounds before the first node dies.

Although this approach is significantly better than the direct transmissions to the BS, there is still some room to save even more energy. The cost of the overhead to form the clusters in LEACH is expensive. In LEACH, in every round, five percent of nodes are cluster-heads and these nodes must broadcast a signal to reach all nodes to determine the members in their clusters. This overhead has been eliminated in the improved version, LEACH-C [14]; otherwise, LEACH-C is identical to LEACH in collection of data in each round. However, several cluster-heads, typically five in a network of 100 nodes, transmit the fused data from the cluster to the distant BS. Further improvement in energy cost for data gathering can be achieved if only one node transmits to the BS per round and if each node transmits only to local neighbors in the data fusion phase. This is exactly what is done in the PEGASIS protocol (defined in Section 4) to obtain an additional factor of two or more improvement compared to LEACH and LEACH-C.

For the 100-node network shown in Fig. 1, we can determine a bound on the maximum number of rounds possible before the first node dies. In each round, every node must transmit their packet and some node or the BS must receive it. So, each node spends two times the energy cost for electronics and some additional cost, depending on how far a node transmits its data. Since at least one node must transmit the fused message to the BS in each round, on the average each node must incur this cost at least once every 100 rounds. With the energy cost parameters and the dimensions of the playing field in Fig. 1 with 100 nodes and 2,000 bit messages, we can calculate the maximum rounds possible. The energy spent in each node for 100 rounds is about 100*0.0002 joules for the electronics and at least 0.002 joules for one message transmission to the BS. With an initial energy in each node of .25 joules, the maximum number of rounds possible before a node dies is given by: $(100 \times 0.25)/0.022 \approx 1,100$.

The actual number of rounds achievable before a node dies will be less since we did not account for the energy spent in the variable part of transmissions, which depends on the distance of transmission and the cost for data fusion. Since each node needs to transmit its data at least to its

closest neighbor, there can be about five to 10 percent more energy cost per round. The exact value clearly depends on the distribution of nodes in the network. Therefore, the upper bound will likely be less than 1,000 rounds. The PEGASIS protocol achieves about 800 rounds, which will likely be within 15-20 percent of this upper bounds, and therefore can be claimed to be near optimal. The following section presents the details of the PEGASIS protocol.

4 PEGASIS: POWER-EFFICIENT GATHERING IN SENSOR INFORMATION SYSTEMS

The main idea in PEGASIS is for each node to receive from and transmit to close neighbors and take turns being the leader for transmission to the BS. This approach will distribute the energy load evenly among the sensor nodes in the network. We initially place the nodes randomly in the playing field and, therefore, the ith node is at a random location. The nodes will be organized to form a chain, which can either be computed in a centralized manner by the BS and broadcast to all nodes or accomplished by the sensor nodes themselves using a greedy algorithm. If the chain is computed by the sensor nodes, they can first get all sensor nodes location data and locally compute the chain using the same greedy algorithm. Since all nodes have the same location data and run the same algorithm, they will all produce the same result. We used random 50, 100, and 200-node networks for our simulations with similar parameters used in [12]. Since this chain computation is done once, followed by many rounds of data communication (typically, several hundred rounds, as shown later), the energy cost in this overhead is small compared to the energy spent in the data collection phase. Therefore, in comparing various schemes, we only consider the energy cost for data collection, fusion, and transmission to the BS and evaluate when the first node dies. With our assumption of no mobility, there will be no change in the chain in the case of PEGASIS and no change in clusters in LEACH-C until the first node dies.

For constructing the chain, we assumed that all nodes have global knowledge of the network and employed the greedy algorithm. We could have constructed a loop. However, to ensure that all nodes have close neighbors is difficult as this problem is similar to the traveling salesman problem. The greedy approach to constructing the chain works well and this is done before the first round of communication. To construct the chain, we start with the furthest node from the BS (select a node randomly if there is a tie). The closest neighbor to this node will be the next node on the chain. Successive neighbors are selected in this manner among unvisited nodes (with ties broken arbitrarily) to form the greedy chain. We begin with the farthest node in order to make sure that nodes farther from the BS have close neighbors as, in the greedy algorithm, the neighbor distances will increase gradually since nodes already on the chain cannot be revisited. Fig. 2 shows node $c0$ connecting to node $c1$, node $c1$ connecting to node $c2$, and node $c2$ connecting to node $c3$, in that order. When a node dies, the chain is reconstructed in the same manner to bypass the dead node.

For gathering data from sensor nodes in each round, each node receives data from one neighbor, fuses the data with its own, and transmits to the other neighbor on the chain. Note that node i will be in some random position j on the chain. Nodes take turns transmitting to the BS and we will use node number i mod N (N represents the number of
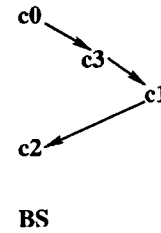

Fig. 2. Chain construction using the greedy algorithm.

nodes) to transmit to the BS in round i. Thus, the leader in each round of communication will be at a random position on the chain, which is important for nodes to die at random locations. The idea of nodes dying at random places is to make the sensor network robust to failures.

Each round of data collection can be initiated by the BS with a beacon signal which will synchronize all sensor nodes. Since all nodes know their positions on the chain, we can employ a time slot approach for transmitting data. In the ith round of data collection, node $c(i-1)$ will be the leader. The end node $c0$ will transmit its data to node $c1$ in slot one, $c1$ fuses and transmits data in slot two, and so on until the leader node is reached. In subsequent slots, data transmissions happen from the node $c(N-1)$ and move toward the leader node from the right end of the chain. Finally, in the Nth slot, the leader transmits data to the BS.

Alternatively, in a given round, we can use a simple control token passing approach initiated by the leader to start the data transmission from the ends of the chain. The cost is very small since the token size is very small. In Fig. 3, node $c2$ is the leader and it will pass the token along the chain first to node $c0$. Node $c0$ will pass its data toward node $c2$. After node $c2$ receives data from node $c1$, it will pass the token to node $c4$, and node $c4$ will pass its data towards node $c2$ with data fusion taking place along the chain.

PEGASIS performs data fusion at every node except the end nodes in the chain. Each node will fuse its neighbor's data with its own to generate a single packet of the same length and then transmit that to its other neighbor (if it has two neighbors). In the above example, node $c0$ will transmit its data to node $c1$. Node $c1$ fuses node $c0$'s data with its own and then transmits to the leader. After node $c2$ passes the token to node $c4$, node $c4$ transmits its data to node $c3$. Node $c3$ fuses node $c4$'s data with its own and then transmits to the leader. Node $c2$ waits to receive data from both neighbors and then fuses its data with its neighbors' data. Finally, node $c2$ transmits one message to the BS. Thus, in PEGASIS, each node, except the two end nodes and the leader node, will receive and transmit one data packet in each round and be the leader once every N rounds. In addition, nodes receive and transmit very small control token packets.

With our simulation experiments, we found that the greedy chain construction performs well with different size networks and random node placements. In constructing the chain, it is possible that some nodes may have relatively distant neighbors along the chain. Such nodes will dissipate more energy in each round compared to other sensors. We improved the performance of PEGASIS by not allowing such nodes to become leaders. We accomplished this by setting a threshold on neighbor distance to be leaders. We may be able to slightly improve the performance of PEGASIS further by applying a threshold adaptive to the

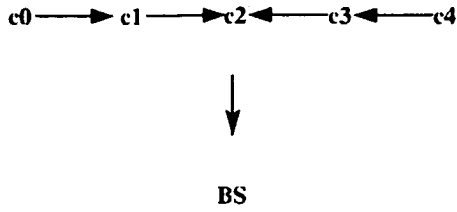c0 ——► c1 ——►c2◄——c3◄——c4

↓

BS

Fig. 3. Token passing approach.

remaining energy levels in nodes. Whenever a node dies, the chain will be reconstructed and the threshold can be changed to determine which nodes can be leaders.

PEGASIS protocol improves on LEACH by saving energy in several stages. First, in the local gathering, the distances that most of the nodes transmit are much less compared to transmitting to a cluster-head in LEACH. Second, the amount of data for the leader to receive is at most two messages instead of 20 (20 nodes per cluster in LEACH for a 100-node network). Finally, only one node transmits to the BS in each round of communication.

5 *Energy* × *Delay* ANALYSIS FOR DATA GATHERING

In this section, we will analyze the *energy* × *delay* cost per round for data gathering from a sensor web to the distant BS. The delay cost can be calculated as units of time. On a 2Mbps link, a 2,000 bit message can be transmitted in $1ms$. Therefore, each unit of delay will correspond to about $1ms$ time for the case of a single channel and non-CDMA sensor nodes. The actual delay value will be different with CDMA nodes, depending on the effective data rate. For each of the systems, we assume that the delay is one unit for each 2,000 bit message transmitted.

The *energy* × *delay* cost for data gathering in a network of N nodes will be different for the schemes considered in this paper and will depend on the node distribution in the playing field. Consider an example network where the N nodes are along a straight line with equal distance of d between each pair of nodes and the BS is a far distance from all nodes. The direct transmission to the BS scheme will require high energy cost and the delay will be N as nodes transmit to the BS sequentially. The PEGASIS scheme forms a chain among the sensor nodes so that each node will receive from and transmit to a close neighbor. For this linear network with equally spaced nodes, the energy cost in PEGASIS is minimized and the variable cost is proportional to $N \times d^2$ and the delay will be N units. Therefore, the *energy* × *delay* cost will be $N^2 \times d^2$.

In the binary scheme with perfect parallel transmission of data, there will be $N/2$ nodes transmitting data to their neighbors at distance d in the lowest level. The nodes that receive data will fuse the data with their own data and will be active in the next level of the tree. Next, $N/4$ nodes will transmit data to their neighbors at a distance $2d$ and this procedure continues until a single node finally transmits the combined message to the BS. Thus, for the binary scheme, the energy cost will be:

$$N/2 \times d^2 + N/4 \times (2d)^2 + N/8 \times (4d)^2 + \ldots + 1 \times (N/2 * d)^2$$

since the distance doubles as we go up the hierarchy. In addition, there will be a single transmission to the BS and the energy cost depends on the distance to the BS. Without including this additional cost by simplifying the above

expression we get for the energy cost for the binary scheme as:

$$N/2 \times d^2 \times (1 + 2 + 4 + \ldots + N/2),$$

which equals

$$N(N - 1)/2 \times d^2.$$

With the additional transmission to the BS, N we can approximate the total energy cost for the binary scheme to be:

$$N^2/2 \times d^2.$$

With the delay cost of about $\log N$ units, the *energy* × *delay* cost for the binary scheme is $N^2/2 \times d^2 \times \log N$. Therefore, for this linear network, the binary scheme will be more expensive than PEGASIS in terms of *energy* × *delay*. For random distribution of nodes in a rectangular playing field, the distances do not double as we go up the hierarchy in the binary scheme and the reduced delay will help reduce the *energy* × *delay* cost. It is difficult to analyze this cost for randomly distributed nodes and we will use simulations to evaluate this cost.

For the rest of the analysis, we assume 50, 100, and 200-node sensor networks in a square field with the BS located far away. In this scenario, energy costs can be reduced if the data is gathered locally among the sensor nodes and only a few nodes transmit the fused data to the BS. This is the approach taken in LEACH [12], where clusters are formed dynamically in each round and cluster-heads (leaders for each cluster) gather data locally and then transmit to the BS. Cluster-heads are chosen randomly, but all nodes have a chance to become a cluster-head in LEACH to balance the energy spent per round by each sensor node. Nodes are able to transmit simultaneously to their cluster-heads using CDMA. For a 100-node network in a $50m \times 50m$ field with the BS located at (25, 150), which is at least 100 meters from the closest node, LEACH reduces the *energy* × *delay* cost compared to the direct scheme. For the linear network of N nodes that are equally spaced, LEACH will have slightly higher energy compared to PEGASIS due to the cluster-heads transmissions to the BS and a delay of roughly N/c, where c is the number of clusters. With five clusters suggested in [12], the *energy* × *delay* for LEACH will be lower than for PEGASIS for a $50m \times 50m$ network. However, for a $100m \times 100m$ network, the *energy* × *delay* for LEACH will be higher than for PEGASIS since PEGASIS achieves increased energy savings with more sparse networks.

The next two sections present protocols that are designed to minimize the *energy* × *delay* metric.

6 A CHAIN-BASED BINARY APPROACH USING CDMA CAPABLE SENSOR NODES

First, we consider a sensor network with nodes capable of CDMA communication. With this CDMA system, it is possible for node pairs that communicate to use distinct codes to minimize radio interference. Thus, parallel communication is possible among 50 pairs for a 100-node network. In order to minimize the delay, we will combine data using as many pairs as possible in each level, which results in a hierarchy of $\lceil \log N \rceil$ levels. At the lowest level, we will construct a linear chain among all the nodes, as was

BS
↑
c3

c3 ◄——— c7

c1 ———► c3   c5 ———► c7

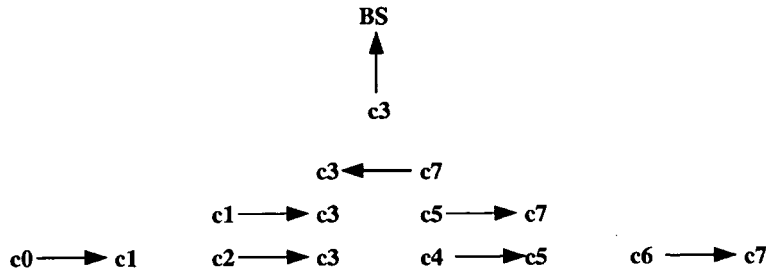c0 ———► c1   c2 ———► c3   c4 ———►c5   c6 ———► c7

Fig. 4. Data gathering in a chain-based binary scheme.

done in PEGASIS, so that adjacent nodes on the chain are nearby. For constructing the chain, we assume that all nodes have global knowledge of the network and employ the greedy algorithm described in Section 4.

For gathering data in each round, each node transmits to a close neighbor in a given level of the hierarchy. This occurs at every level in the hierarchy, but the only difference is that the nodes that are receiving at each level are the only nodes that will be active in the next level. Finally, at the top level, the only node remaining will be the leader and the leader will transmit the k bit message to the BS. Note that node i will be in some random position j on the chain. Nodes take turns transmitting to the BS and we will use node number i mod N (N represents the number of nodes) to transmit to the BS in round i. In Fig. 4, for round three (first round is round zero), node $c3$ is the leader. Since, node $c3$ is in position 3 (counting from 0) on the chain, all nodes in an even position will send to their right neighbor. Now, at the next level, node $c3$ is still in an odd position, so, again, all nodes in an even position will fuse their data with its received data and send to their right. At the third level, node $c3$ is not in an odd position, so node $c7$ will fuse its data and transmit to $c3$. Finally, node $c3$ will combine its current data with that received from $c7$ and transmit the message to BS.

The chain-based binary scheme performs data fusion at every node that is transmitting except the end nodes in each level. Each node will fuse its neighbor's data with its own to generate a single packet of the same length and then transmit that to the next node. In the above example, node $c0$ will pass its data to node $c1$. Node $c1$ fuses node $c0$'s data with its own and then transmits to node $c3$ in the next level. In our simulations, we ensure that each node performs an equal number of sends and receives after N rounds of communication and each node transmits to the BS in one of N rounds. We then calculate the average energy cost per round, while the delay cost is the same for each round. We compute the average energy × delay cost over a number of different node distributions. Experimental results are presented in detail in Section 8.

The chain-based binary scheme improves on LEACH and LEACH-C by saving energy and delay in several stages. At the lower levels, nodes are transmitting at shorter distances compared to nodes transmitting to a cluster-head in the LEACH protocol and only one node transmits to the BS in each round of communication. By allowing nodes to transmit simultaneously, the delay cost for the binary scheme decreases from that of LEACH by a factor of about three. While, in LEACH and LEACH-C, only five groups can transmit simultaneously for a 100-node network, here, at each level, we have more nodes transmitting simultaneously. At each level of the binary scheme, transmissions

are simultaneous, making the total delay $\lceil \log N \rceil + 1$, including the transmission to the BS. In LEACH and LEACH-C, the delay for 100-node networks will be 27 units. The delay for all nodes to transmit to the cluster-head is the max number of nodes in any of the five clusters. If all the clusters are of the same size, then the delay would be 19. Then, all five cluster-heads must take turns to transmit to the BS, making that a total of 24. For overhead calculations, we have one unit of delay for cluster formation, one unit of delay for all nodes to broadcast to the cluster-head its presence in that cluster, and, finally, one unit of delay for the cluster-head to broadcast a schedule sequence to the nodes so that all nodes within a cluster know when to transmit their data to the cluster-head.

7 A CHAIN-BASED THREE LEVEL SCHEME WITHOUT CDMA CAPABLE SENSOR NODES

CDMA may not be applicable for all sensor networks as these nodes can be expensive. Therefore, we need a protocol that will achieve a minimal energy × delay with non-CDMA nodes. It will not be possible to use the binary scheme in this case as the interference will be too much at lower levels. We either have to increase the energy cost significantly or take more time steps at lower levels of the hierarchy, both of which will lead to much higher energy × delay cost. Therefore, in order to improve energy × delay, we need a protocol that allows simultaneous transmissions that are far apart to minimize interference while achieving reasonable delay cost. Based on our experiments, we suggest the chain-based 3-level scheme for data gathering in sensor networks with non-CDMA nodes.

Also, in the 3-level scheme, we start with the linear chain among all the nodes and divide them into G groups, with each group having N/G successive nodes of the chain. Therefore, we will have G groups of N/G nodes. One node from each group will be active in the second level and, thus, there will be G nodes. These G nodes in the second level are divided into two groups of successive nodes in order to maintain only three levels in the hierarchy. G is calculated based on the number of nodes and the size of the network. For a $100m \times 100m$ network, we found that, when G is equal to 10, we get the best balance for energy and delay. In a 100-node network, therefore, only 10 simultaneous transmissions take place at the same time and data fusion takes place at each node (except the end nodes in each level). The transmissions are also far enough apart that there is minimal interference and we can still maintain low energy costs at each level in the hierarchy while maintaining a low delay. Fig. 5 shows an example of this scheme with 100 nodes. We will have a different leader in each round transmit to the BS to evenly distribute the workload

BS

$\uparrow$

c18

c18 ← c68

c8 → c18 ← c28 ← c38 ← c48   c58 → c68 ← c78 ← c88 ← c98

c0 → c1 → c2 → c3 → c4 → c5 → c6 → c7 → c8 ← c9 → c10 → c11 ... c18 ← c19 ... c90 → c91 ... c98 ← c99
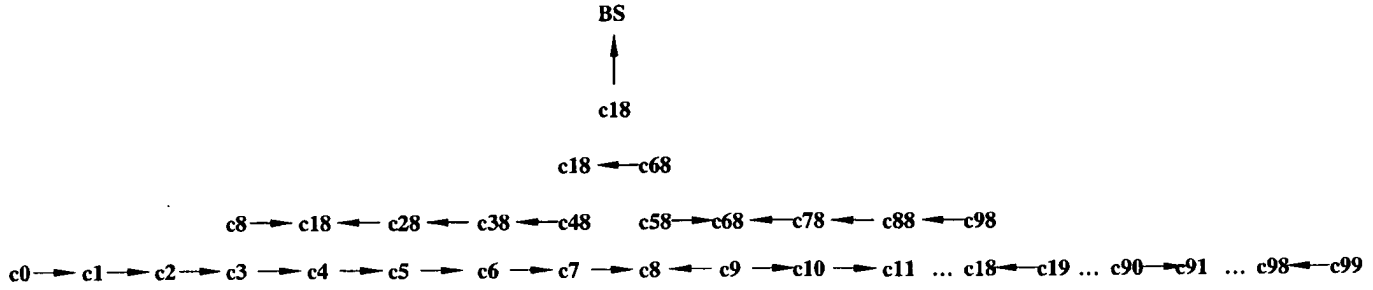
Fig. 5. Chain-based 3-level scheme for a sensor network with non-CDMA nodes.

among the sensor nodes. As before, we will use node i along the chain to be the leader in the ith round of communication. We find the index i within a group which will represent the leader position modulo N/G.

In Fig. 5, node $c18$ is our leader. Then all nodes will send their data in the direction of index 8 within their group since 18 modulo 10 is 8. The delay at the first level is nine units. Then the second level will contain nodes $c8, c18, c28 \ldots c98$. These 10 nodes will be divided into two groups. If we have more levels in the hierarchy, then distances between nodes become further apart, causing higher energy costs. By experimentation, for the networks under consideration, having three levels gives us the best balance of energy and delay. Since the leader position is 18, all nodes that are in the first group will send down the chain 10 positions from its own position on the chain. So, node $c48$ will send to node $c38$, and node $c38$ will send to node $c28$ and so on. Since node $c8$'s position is less than node $c18$'s, node $c8$ will transmit to a position that is N/G greater than its own. In group two, nodes know in which direction to send the data using the leader position $N/2$. So, here, the nodes in group two would send in the direction of node $c68$ in the same manner as in group one. This gives us a delay of four units for the second level. In the third level, node $c68$ transmits to node $c18$, who is our leader, and then, finally, node $c18$ transmits the combined packet to the BS, giving us a total delay of 15 units. The transmission schedule can be programmed once at the beginning so that all nodes know where to send data in each round of communication.

8 EXPERIMENTAL RESULTS

This section presents the performance analysis of the different protocols using simulation programs written in C programming language. We used several simulation parameter variations to test our schemes. The network dimensions studied were $50m \times 50m$ and $100m \times 100m$. The BS locations were varied at (50, 150), (50, 200), and (50, 300). The packet sizes considered were 2,000, 10,000, and 20,000 bits. The number of nodes were varied as 50, 100, and 200 to test for dense and sparse networks. Extensive simulations were run to determine the optimal number of clusters to use when the number of nodes varied for the LEACH protocol. The LEACH protocol uses five clusters for a 100-node network. We found that, for a 200-node network, five clusters were optimal, and, for a 50-node network, two clusters were optimal.

8.1 Comparison of LEACH and PEGASIS Using the Energy Metric

For this experiment, the metric studied was the number of rounds of communication achieved when 1 percent,

25 percent, 50 percent, and 100 percent of the nodes die using direct transmission, LEACH, and PEGASIS. Each node is assumed to have the same initial energy level of 0.25J. Once a node dies due to battery power depletion, it is not recharged for the rest of the simulation. LEACH-C improves upon LEACH by about 40 percent due to the centralized computation by the BS to find better clusters [14]. Therefore, as stated before, in the rest of this section, we present our comparison results only with LEACH. The performance improvements will be correspondingly lesser compared to LEACH-C to the extent LEACH-C improves upon LEACH, which is about 20 percent to 40 percent, depending on network parameters [14].

Fig. 6 shows the number or rounds until 1 percent, 25 percent, 50 percent, and 100 percent nodes die for a $50m \times 50m$ network. PEGASIS is approximately two times better than LEACH in all cases for a $50m \times 50m$ network. The overhead energy cost in forming clusters in LEACH or chain in PEGASIS are similar. It may be more useful to compute this centrally in the BS, which doesn't have an energy limitation. The improvements in PEGASIS come due to fewer nodes transmitting data to BS in each round compared to LEACH and its variants.

The next set of experiments were conducted for a $100m \times 100m$ network. Fig. 7 shows the number of rounds completed for the same percentages of node deaths with different locations of the BS. The BS locations are at (50, 150), (50, 200), and (50, 300).

The simulation results show that PEGASIS achieves:

- approximately two times the number of rounds compared to LEACH when 1 percent, 25 percent, 50 percent, and 100 percent of nodes die for a $50m \times 50m$ network,
- approximately three times the number of rounds compared to LEACH when 1 percent, 25 percent, 50 percent, and 100 percent nodes die for a $100m \times 100m$ network,
- balanced energy dissipation among the sensor nodes to have full use of the complete sensor network,
- near-optimal performance.

However, there are some rare cases when the first node death occurs with PEGASIS slightly earlier in comparison to LEACH, as shown in Fig. 7a. This is due to the greedy chain construction procedure used, where a node may have a local neighbor very far away and thus will deplete energy more quickly and die first. This happens only for some distribution of nodes and an approach to ensure that PEGASIS always performs best before the first node death occurs is to construct a chain so that all nodes have relatively close neighbors. To construct such a chain
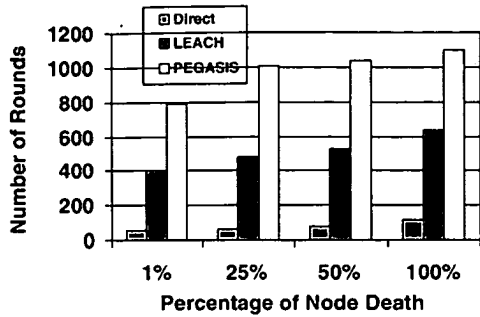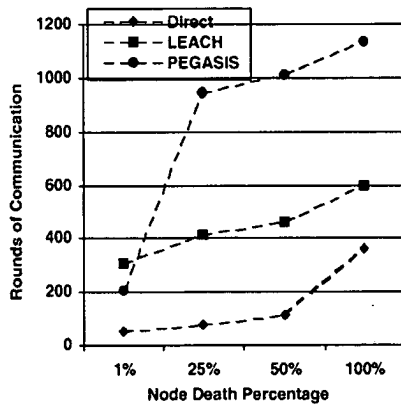
Fig. 6. Performance results for a $50m \times 50m$ network, BS location at (25, 150), and 100 nodes.

requires the use of the global knowledge of all node positions to pick suitable neighbors and minimize the maximum neighbor distance. This problem is related to the traveling salesman problem of minimizing the total length of the loop (chain), which is known to be intractable. Heuristic algorithms to solve this problem can be expensive compared to the simple scheme used in PEGASIS and the advantages are minimal as PEGASIS is nearly optimal in terms of rounds achievable when a larger percentage of nodes die.

## 8.2 Comparison of All Schemes Using the *Energy* × *Delay* **Metric**

To evaluate the performance of the chain-based binary scheme and the chain-based 3-level scheme, we simulated direct transmission, PEGASIS, LEACH, and the two new schemes using several random 50, 100, and 200-node networks with CDMA nodes and non-CDMA nodes. We used the same simulation parameters as described above for evaluating PEGASIS. However, instead of running the simulations for percentage of node deaths, we ran the simulations for enough rounds in all the schemes so that all N nodes had a chance to become leader only once. Since different schemes have to run for a different number of rounds before every node has a chance to become leader only once, it does not make sense to compare the number of rounds before nodes die. By doing this, we can compare the average energy costs per round for all the schemes fairly. We then used these costs to determine the average energy cost per round of data gathering for several different topologies. To calculate the *energy* × *delay* for these schemes, we multiply the average energy cost per round to the unit delay for the scheme. In both CDMA and non-CDMA systems, we included the interference costs when there are simultaneous transmissions to ensure that the same SNR of 10 dB is maintained as with single transmission. For the 3-level scheme, we evaluated the number of

(a)
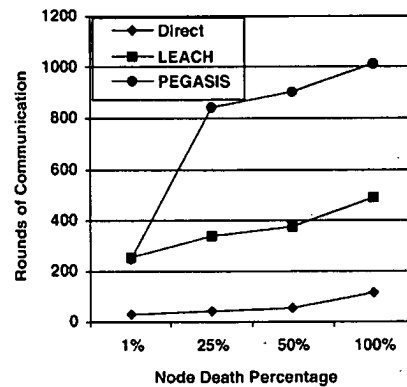
(b)
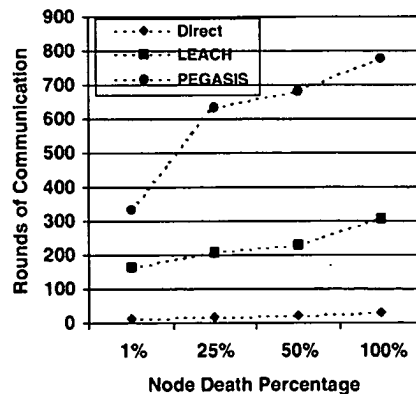
(c)

Fig. 7. Performance results for a $100m \times 100m$ network with the BS location at (a) pos = (50, 150), (b) pos = (50, 200), and (c) pos = (50, 300). The packet size is 2,000 bits and the number of nodes is 100.

TABLE 1
Energy × delay Cost for Direct, PEGASIS, LEACH,
Chain-Based Binary Scheme and the Chain-Based Three Level Scheme

Protocol	Energy		Delay	Energy × Delay	
	D = 50	D = 100	D = 50 & D = 100	D = 50	D = 100
Direct (both systems)	0.3299	1.2805	100	32.9938	128.0459
PEGASIS (both systems)	0.0240	0.0361	100	2.4008	3.6107
LEACH (CDMA nodes)	0.0797	0.2048	27	2.1518	5.5292
Chain-based binary (CDMA nodes)	0.0318	0.0559	8	0.2547	0.4516
Chain-based 3 level (non-CDMA nodes)	0.0358	0.0583	15	0.5365	0.8743

These results are for a 50m × 50m and 100m × 100m network where D equals the dimension of a D × D network.

9 CONCLUSIONS AND FUTURE WORK

groups for the first level when the number of nodes change in the network to guarantee the optimal *energy × delay*. We found that, for 50-node, 100-node, and 200-node networks, dividing the nodes into 10 groups gave us the optimal *energy × delay*.

Table 1 gives the results for energy cost, delay cost, and *energy × delay* cost for direct, PEGASIS, LEACH, the chain-based binary scheme, and the chain-based 3-level scheme.

Fig. 8 shows the results for the five schemes based on different BS locations. *Energy × delay* is higher for all schemes as the BS moves farther away from the nodes. Fig. 9 shows the results for the five schemes based on different packet sizes. As expected, *energy × delay* increases with the packet size. Fig. 10 shows that, as the number of nodes increase, *energy × delay* becomes greater for all schemes. For all these figures, the binary scheme performs the best however if sensors are not CDMA capable, then the 3-level scheme is the best.

The simulation results show that:

- The chain-based binary scheme is approximately eight times better than LEACH and 130 times better than direct for a 50m × 50m network in terms of *energy × delay* for sensor networks with CDMA nodes.
- The chain-based binary scheme is approximately five to 13 times better than LEACH and 80 or more times better than the direct scheme for a 100m × 100m network in terms of *energy × delay* for sensor networks with CDMA nodes.
- The chain-based three level scheme is approximately four times better than PEGASIS and 60 times better than direct for a 50m × 50m network in terms of *energy × delay* for sensor networks with non-CDMA nodes.
- The chain-based 3-level scheme is approximately three to five times better than PEGASIS and up to 140 times better than direct for a 100m × 100m network in terms of *energy × delay* for sensor networks with non-CDMA nodes.
- The chain-based schemes show a more balanced energy dissipation among the sensor nodes to have full use of the complete sensor network.

In this paper, we describe three new protocols for wireless sensor networks. One of these protocols, PEGASIS, is a greedy chain protocol that is near optimal for a data-gathering problem in sensor networks. PEGASIS out-performs LEACH by eliminating the overhead of dynamic cluster formation, minimizing the distance nonleader nodes must transmit, limiting the number of transmissions and receptions among all nodes, and using only one transmission to the BS per round. Nodes take turns to transmit the fused data to the BS to balance the energy depletion in the network and preserve the robustness of the sensor web as nodes die at random locations. Distributing the energy load among the nodes increases the lifetime and quality of the network. Our simulations show that PEGASIS performs better than LEACH by about 100 to 200 percent when 1 percent, 25 percent, 50 percent, and 100 percent of nodes die for different network sizes and topologies. The improvements will be slightly lesser compared to LEACH-C, which doesn't have the cluster formation over-head in each round. PEGASIS shows an even further improvement as the size of the network increases.

The other two protocols described in this paper that reduce the energy as well as delay for data gathering in sensor networks are a chain-based binary scheme for sensor networks with CDMA nodes and a chain-based 3-level scheme for sensor networks with non-CDMA nodes. The
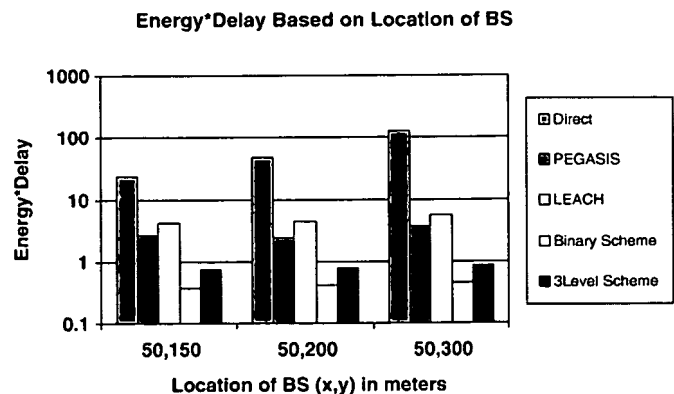
Energy*Delay Based on Location of BS

Fig. 8. Performance results for a 100m × 100m network with BS locations at (50, 150), (50, 200), and (50, 300). The packet size is 2,000 bits and the number of nodes in the network is 100.
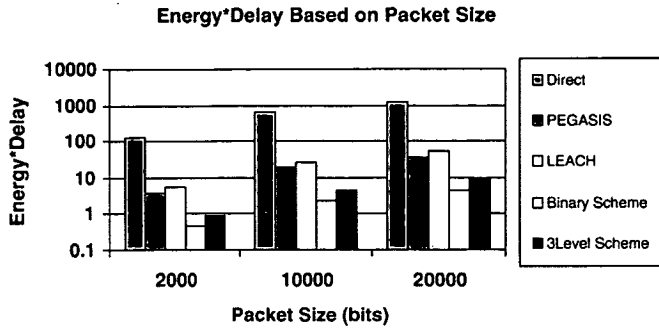
Fig. 9. Performance results for a $100m \times 100m$ network with BS location at (50, 300). The packets sizes vary at 2,000, 10,000, and 20,000 bits and the number of nodes in the network is 100.
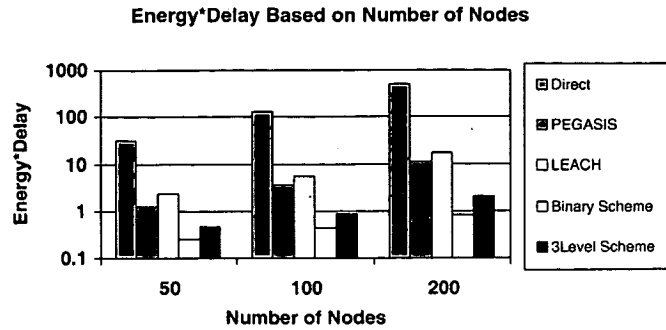

Fig. 10. Performance results for a $100m \times 100m$ network with the BS location at (50, 300). The packet size is 2,000 bits and the number of nodes vary at 50, 100, and 200.

binary scheme performs better than direct, PEGASIS, and LEACH. It performs better than LEACH by a factor of about eight, about 10 times better than PEGASIS, and more than 100 times better when compared to the direct scheme. In these experiments, the interfering transmissions contributions are assumed to be 1/128 the value of their transmission energy. With non-CDMA nodes, the interfering energy is the amount received from unintended transmissions. The chain-based 3-level scheme with non-CDMA outperforms PEGASIS by a factor of four and is better than direct by a factor of 60. The scheme outperforms PEGASIS by dividing the chain in "groups" and allowing simultaneous transmissions among pairs in different groups. While energy is still minimal, the delay is decreased from 100 units to 15 units.

It is not clear as to what is the optimal scheme for optimizing $energy \times delay$ is in a sensor network. Since the energy costs of transmissions depend on the spatial distribution of nodes, there may not be a single scheme that is optimal for all sizes of the network. Our preliminary experimental results indicate that, for all small networks, the binary scheme performs best as minimizing delay achieves best result for $energy \times delay$. With larger networks, we expect that nodes in the higher levels of the hierarchy will be far apart and it is possible that a different multilevel scheme may outperform the binary scheme. When using non-CDMA nodes, interference effects can be reduced by carefully scheduling simultaneous transmissions. Since there is an exponential number of possible schedules, it is intractable to determine the optimal scheduling to minimize $energy \times delay$ cost. A practical scheme to employ will depend on the size of the playing field and the distribution of nodes in the field.

In order to validate our assumptions, more detailed models and a network simulator, such as ns-2, need to be used for detailed evaluations. Based on our C simulations, we expect that PEGASIS will outperform LEACH and its variants and direct protocols in terms of system lifetime and the quality of the network for minimizing energy. We also expect that the binary chain-based scheme and the 3-level chain-based scheme will outperform direct, LEACH and its variants, and PEGASIS in terms of $energy \times delay$. We also restricted our discussions to the d^2 model for energy dissipation for wireless communications in this paper. In our future work, we will consider higher order energy dissipation models and develop schemes to minimize energy and $energy \times delay$ costs for this type of data gathering and other applications in sensor networks.

REFERENCES

[1] "The Ultimate Sting: Bees the Buzz in Landmine Detection,"http://www.pnl.gov/news/1999/99-13.htm. Apr. 2001.
[2] "Bluetooth Initiative,"http://www.bluetooth.com, 1999.
[3] N. Bambos, "Toward Power-Sensitive Network Architectures in Wireless Communications: Concepts, Issues, and Design Aspects," IEEE Personal Comm., pp. 50-59, June 1998.
[4] A. Chandrakasan, R. Amirtharajah, S.-H. Cho, J. Goodman, G. Konduri, J. Kulik, W. Rabiner, and A. Wang, "Design Considerations for Distributed Microsensor Networks," Proc. IEEE Custom Integrated Circuits Conference (CICC), pp. 279-286, 1999.
[5] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "Span: An Energy-Efficient Coordination Algorithm for Topology Maintenance in Ad Hoc Wireless Networks," Proc. SIGMOBILE, July 2001.
[6] L. Clare, G. Pottie, and J. Agre, "Self-Organizing Distributed Sensor Networks," Proc. SPIE Conf. Unattended Ground Sensor Technologies and Applications, pp. 229-237, 1999.
[7] "COTS Dust,"http://www-bsac.eecs.berkeley.edu/~shollar/macro_motes/macromotes.html. Jan. 2001.
[8] M. Dong, K. Yung, and W. Kaiser, "Low Power Signal Processing Architectures for Network Microsensors," Proc. Int'l Symp. Low Power Electronics and Design, pp. 173-177, 1997.
[9] D. Dudgeon and R. Mersereau, Multidimensional Signal Processing. Prentice Hall, 1984.
[10] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next Century Challenges: Scalable Coordination in Sensor Networks," Proc. ACM/IEEE Mobicom, Aug. 1999.
[11] M. Ettus, "System Capacity, Latency, and Power Consumption in Multihop-Routed SS-CDMA Wireless Networks," Proc. Radio and Wireless Conference (RAWCON), 1998.
[12] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," Proc. Hawaii Conf. System Sciences, Jan. 2000.
[13] W. Heinzelman, A. Sinha, A. Wang, and A. Chandrakasan, "Energy-Scalable Algorithms and Protocols for Wireless Microsensor Networks," Proc. Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP), June 2000.
[14] W. Heinzelman, "Application-Specific Protocol Architectures for Wireless Networks," PhD thesis, Massachusetts Inst. of Technology, June 2000.
[15] L. Klein, Sensor and Data Fusion Concepts and Applications. SPIE Optical Eng. Press, 1993.

[16] J. Kulik, W. Rabiner, and H. Balakrishnan, "Adaptive Protocols for Information Dissemination in Wireless Sensor Networks," *Proc. ACM/IEEE Mobicom*, Aug. 1999.

[17] S. Lindsey, K.M. Sivalingam, and C.S. Raghavendra, "Power-Aware Routing and MAC Protocols for Wireless and Mobile Networks," *Wiley Handbook on Wireless Networks and Mobile Computing*, pp. 407-423, John Wiley & Sons, 2002.

[18] K. Nakano and S. Olariu, "Energy-Efficient Initialization Protocols for Single-Hop Radio Networks with No Collision Detection," *IEEE Trans. Parallel and Distributed Systems* vol. 11, no. 8, Aug. 2000.

[19] K. Nakano and S. Olariu, "Randomized Initialization Protocols for Ad Hoc Networks," *IEEE Trans. Parallel and Distributed Systems*, vol. 11, no. 7, July 2000.

[20] R. Pichna and Q. Wang, "Power Control," *The Mobile Comm. Handbook*, pp. 370-380, CRC Press, 1996.

[21] T.S. Rappaport, *Wireless Communications*, Prentice-Hall, 1996.

[22] R. Ramanathan and R. Hain, "Topology Control of Multihop Wireless Networks Using Transmit Power Adjustment," *Proc. IEEE Infocom*, Mar. 2000.

[23] E. Shih, S.-H. Cho, N. Ickes, R. Min, A. Sinha, A. Wang, and A. Chandrakasan, "Physical Layer Driven Algorithm and Protocol Design for Energy-Efficient Wireless Sensor Networks," *Proc. SIGMOBILE*, July 2001.

[24] S. Singh and C. S. Raghavendra, "PAMAS: Power Aware Multi-Access Protocol with Signalling for Ad Hoc Networks" *ACM Computer Comm. Rev.*, July 1998.

[25] K. Sivalingam, J. Chen, P. Agrawal, and M. Srivastava, "Design and Analysis of Low-power Access Protocols for Wireless and Mobile ATM Networks," *ACM/Baltzer Wireless Networks*, vol. 6, pp. 73-87, Feb. 2000.

[26] R. Steele, *Mobile Radio Communications*, London Pentech Press, 1992.

[27] M. Stemm, P. Gauthier, D. Harada, and R. Katz, "Reducing Power Consumption of Network Interfaces in Hand-Held Devices," *Proc. Thirdrd Int'l Workshop Mobile Multimedia Comm.*, pp. 25-27, Sept. 1996.

[28] J. Wieselthier, G. Nguyen, and A. Ephremides, "On the Construction of Energy-Efficient Broadcast and Multicast Trees in Wireless Networks," *Proc. IEEE Infocom*, Mar. 2000.

[29] *The WINS Project*, http://www.janet.ucla.edu/WINS. Feb. 1999.

[30] Y. Xu, J. Heidemann, and D. Estrin, "Geography-Informed Energy Conservation for Ad Hoc Routing," *Proc. SIGMOBILE*, July 2001.

[31] K. Yao, R. Hudson, C. Reed, D. Chen, and F. Lorenzelli, "Blind Beamforming on a Randomly Distributed Sensor Array System," *Proc. Signal Processing Systems*, 1998.

C.S. Raghavendra received the PhD degree in computer science from the University of California at Los Angeles in 1982. He is a professor of electrical engineering and computer science and director of Computer Engineering Division in the Electrical Engineering Department at the University of Southern California (USC), Los Angeles. Previously, he was a faculty member in the Department of Electrical Engineering-Systems at USC from 1982 to 1992, was the Boeing Chair Professor of Computer Engineering in the School of Electrical Engineering and Computer Science at the Washington State University in Pullman, from 1992 to 1997, and was with The Aerospace Corporation from 1997 to 2001. He is a recipient of the Presidential Young Investigator Award for 1985 and is a fellow of the IEEE.

K.M. Sivalingam received the PhD and MS degrees in computer science from the State University of New York (SUNY), at Buffalo in 1994 and 1990, respectively. He received the BE degree in computer science and engineering in 1988 from Anna University, Madras, India.- While at SUNY Buffalo, he was a presidential fellow from 1988 to 1991. He is currently an associate professor in the School of Electrical Engineering and Computer Science, at Washington State University, Pullman, where he was an assistant professor from 1997 to 2000. Earlier, he was an assistant professor at the University of North Carolina, Greensboro, from 1994 to 1997. He has conducted research at Lucent Technologies' Bell Labs in Murray Hill, New Jersey, and at AT&T Labs in Whippany, New Jersey. His research interests include wireless networks, optical wavelength division multiplexed networks, and performance evaluation. He has served as a guest coeditor for a special issue of the *IEEE Journal on Selected Areas in Communications* on optical WDM networks. He is a corecipient of the Best Paper Award at the IEEE International Conference on Networks 2000 held in Singapore. He has published an edited book on optical WDM networks in 2000. His work has been supported by the US Air Force Office of Scientific Research, Laboratory for Telecommunication Sciences, US National Science Foundation, Cisco, Bellcore, Alcatel, Intel, and Washington Technology Center. He holds three patents in wireless networks and has published several research articles, including more than 20 journal publications. He has served as the technical program co-chair of SPIE/IEEE/ACM OptiComm Conference in Boston, in July 2002 and as the workshop co-chair for the Workshop on Wireless Sensor Networks and Applications (WSNA) to be held in conjunction with ACM MobiCom 2002 in Atlanta, in September 2002. He has also served on several international conference committees in various capacities, including ACM Mobicom 2001, Opticomm 2001, Opticomm 2000, ACM Mobicom 1999, MASCOTS 1999, and IEEE INFOCOM 1997. He is a senior member of the IEEE and a member of ACM.

Stephanie Lindsey received the MS degree in computer science from Washington State University, Pullman, in 2000 and the BSc degree in computer science from the University of North Texas, Denton, in 1999. She is currently employed at Microsoft Corporation, Redmond, Washington. Her research interests include wireless ad hoc and sensor networks.

▷ **For more information on this or any computing topic, please visit our Digital Library at** http://computer.org/publications/dilb.

Performance of Adaptive Antennas in FH-GSM Using Conventional Beamforming

PREBEN E. MOGENSEN, POUL LETH ESPENSEN, PER ZETTERBERG,
KLAUS I. PEDERSEN and FRANK FREDERIKSEN
Center for Personkommunikation, Aalborg University, Fr. Bajers Vej 7A6-2, DK-9920 Aalborg, Denmark
E-mail: pm@cpk.auc.dk

Abstract. This paper presents the performance of adaptive antennas in a 1/3 reuse frequency hopping GSM network using conventional beamforming. It mainly focuses on C/I improvement for the purpose of capacity enhancement. The performance evaluation has been conducted by means of network computer simulations, where measured time-space radio channel impulse responses are applied for the desired user in the network. Measurements with an $M = 8$ element uniform linear array were conducted in the cities of Aarhus, Denmark, and Stockholm, Sweden. The simulated C/I improvement shows an almost $10 * \log_{10}(M)$ behavior for low azimuth spread values. For large values of azimuth spread (relative to the antenna beamwidth), the performance gain is reduced significantly. For an azimuth spread of $10°–12°$, which has been measured in urban macro-cellular environments, the C/I gain for $M = 8$ is reduced to approx. 5.5–7.5 dB (which should be compared to the theoretical value of 9 dB for a point source). The designed DoA algorithm is very robust to co-channel interference and only a small degradation in performance is observed for single element C/I down to approx. –8 dB. We conclude that the designed beamforming implementation facilitates a potential capacity gain of ×3 in a 1/3 reuse FH-GSM network for an array size of $M = 4$–6.

Keywords: adaptive antennas, GSM, capacity.

1. Introduction

In recent years, there has been an enormous interest in adaptive base station antenna technology for GSM (among other 2nd and 3rd generation land mobile communication systems) for both enhanced capacity and coverage range [1, 2, 25]. The theory and the potential performance of adaptive antennas has been described in numerous publications [3–5, 22]. However, due to several practical aspects of adaptive antennas, such as hardware implementation [8] and frequency assignment in a network with only a few adaptive base stations [6], a commercial exploitation of adaptive antennas may likely take a slow evolutionary path [7]. The promising theoretical results still need to be validated by large-scale field trials [9].

The EU funded ACTS project TSUNAMI II [10] was established in 1995 to study the feasibility and cost efficiency of deploying adaptive antennas for 3rd generation mobile communication systems. A major effort of the TSUNAMI II project was put into prototyping and conducting field trials with an adaptive antenna base station for GSM-1800 [25]. Being partner in TSUNAMI II, the Center for Personkommunikation at Aalborg University, Denmark, designed a robust beamforming algorithm to be used for the field trials [18].

This paper concentrates on the design and performance of the proposed beamforming algorithm. There were several reasons for mainly concentrating our effort on *conventional* beamforming: (a) the issue of downlink transmission in a frequency division duplex (FDD)

system and (b) the TSUNAMI II adaptive antenna base station used in the field trial was incapable of performing instantaneous weight calculation, which, for example, is required for Optimum Combining (OC) techniques [23] or Maximal Ratio Combing [12]. The paper is organized as follows. In Section 2, we discuss some GSM system aspects relevant for the design of the beamforming algorithm. Section 3 briefly describes the time-space measurements used for the network simulations, while Section 4 presents the simulated network configuration. The proposed beamforming algorithm is derived in Section 5 and the performance for uplink and downlink are shown in Sections 6 and 7, respectively. Finally, conclusions are drawn about the work in Section 8.

2. GSM Capacity Aspects

It is anticipated that slow Frequency Hopping (FH) is a powerful feature for enhancing the capacity of GSM networks. In fact, most GSM operators see FH as the most promising and cost-efficient solution for enhancing their network capacity [11]. Thanks to channel coding and interleaving in GSM [31], FH provides both frequency and interference diversity [11]. When considering high-capacity GSM networks, it is therefore deemed mandatory to consider adaptive base station antennas along with other optional capacity-enhancing features in GSM, such as: FH, Power Control (PC), and discontinuous transmission (DTX).

In a conventional non-FH GSM network, the minimum frequency reuse distance for a three-sectored base station configuration is $K/n = 4/12$ [11], where K is the cluster size and n is the number of frequencies used within a cluster. The 4/12 frequency reuse is always applicable for the BCCH (BroadCast Channel) carrier, because continuous transmission with full power is mandatory. If using synthesized FH on non-BCCH carriers, the number of hopping frequencies (MA list) can be equal to or higher than the number of base station transceivers (TRXs). The term *fractional loading* is often used for MA lists larger than the number of TRXs. In FH GSM, two parameters can thus be used to control the interference level: the cluster size, K, and the fractional loading, L. Both network simulations and field trial results have verified that a GSM network can operate with a frequency reuse of 1/3 with a fractional loading of approx. 25–30%.

By deploying adaptive antenna base stations in GSM, it was our goal to potentially increase the load of a 1/3 reuse network up to the hard-blocking level of 80–90% load, thus achieving a capacity increase on the order of ×3 by adaptive antennas.[1] However, this potential capacity increase has the condition that a sufficient spatial filtering (beamforming) gain is achieved. By assuming a linear relation between network load and the increase in interference level, the minimum required spatial filtering gain can be estimated to approx. 5 dB. Applying a $10 * \log_{10}(M)$ relationship for the spatial filtering gain for a point source, an array size of $M = 4$ should be sufficient. However, when considering azimuth multipath spread, it was initially analyzed in [18] that 6–8 elements could be necessary in urban environments with a median azimuth spread of 10° or higher. In the following, we mainly limit our study to the case of $M = 8$.

---

[1] In [18], a more detailed analysis of the potential capacity gain, including the aspect of BCCH carrier, trunking efficiency, etc., can be found.

3. Space-Time Channel Sounding

3.1. DESCRIPTION OF THE TESTBED

The TSUNAMI II Stand-alone testbed was built with the purpose of performing time-space channel sounding and verifying beamforming algorithms [14]. The testbed consists of a transmitter (mobile unit) and an adaptive array receiver (base station unit). The base station unit has 9 parallel receiver chains. Eight of the receivers are connected to an 8-element antenna array, while the ninth receiver is connected to a sector reference antenna. The carrier frequency is within the GSM-1800 frequency band. For the results presented here, an 8-element uniform linear array antenna was used. The antenna elements are vertically polarized dipoles mounted in front of a reflecting ground plane; the horizontal element spacing is half a wavelength. The bandwidth was 200 kHz according to the GSM specifications [31] in the narrowband channel sounding mode and 4 MHz according to UMTS in the wideband sounding mode [32].

3.2. DESCRIPTION OF THE NARROWBAND MEASUREMENTS, AARHUS

The measurement data from Aarhus, Denmark, was obtained using the 200 kHz bandwidth. The city of Aarhus is characterized by an irregular street layout and 3–5-story buildings. The adaptive base station antenna array was installed at a hotel on three different floors, corresponding to the heights of 20 m, 26 m, and 32 m. The lowest position almost corresponded to the rooftop level of the surrounding buildings. Several measurement routes were measured, and each route was repeated for the three antenna heights. The simulation results obtained from Aarhus are for a route ranging in distance from 1.5 km to 2 km from the base station and within an azimuth region of +20° to –30° relative to the broadside of the antenna array. We characterize the measurement area as Typical Urban, TU.

3.3. DESCRIPTION OF THE WIDEBAND MEASUREMENTS, STOCKHOLM

The measurements from Stockholm, Sweden, were obtained using the upgraded wideband mode of the testbed. The simulation results presented are from a route ranging in distance from 0.5 km to 1 km from the base station, and a azimuth region ranging from –20° to 0°. The base station antenna height was 21 m and the average building height was only a few meters lower. The measurement route for the analysis was close to a river crossing the city of Stockholm, which gave raise to a two-cluster multipath scenario. The propagation scenario is therefore characterized as Bad Urban, BU. The wideband data from Stockholm was bandpass-filtered in order to comply with the GSM bandwidth. The filtering process makes it possible to extract several narrowband GSM frequency channels and thereby emulate FH.

3.4. CHARACTERIZATION OF AZIMUTH DISPERSION

It was published [15] and also verified [16] that the Power Azimuth Spectrum (PAS) for a macro-cell base station can be well described by a Laplacian function for *Typical Urban* (TU) and *Rural Area* (RA):

$$P(\phi) = \exp\left[-\sqrt{2}\frac{|\phi|}{\sigma}\right] \quad \phi \in [-\pi, +\pi[, \tag{1}$$
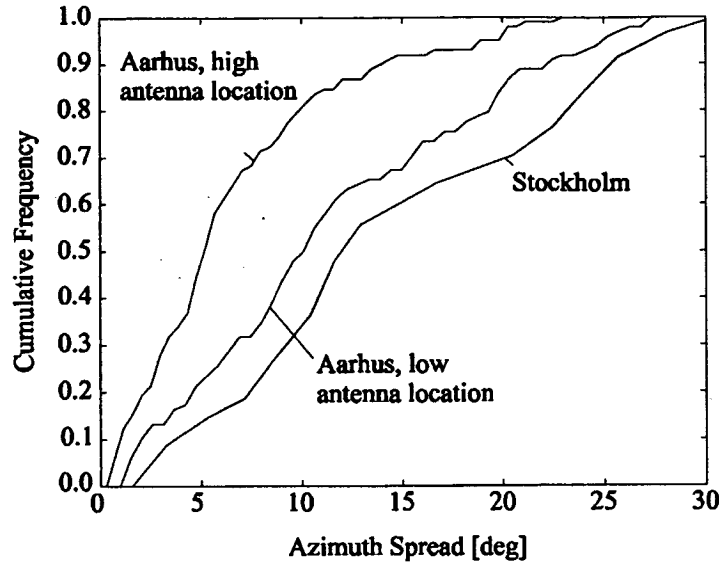
Figure 1. Distribution of azimuth spread for the base station antenna heights of 32 m (high) and 20 m (low) in Aarhus and 21 m in Stockholm [27].

where ϕ is azimuth and σ is the Azimuth Dispersion (AS). For the Bad Urban (BU) case of Stockholm, the PAS could be well modelled by two clusters with different mean azimuths, where each cluster is described by a Laplacian function [27].

Figure 1 shows the cumulative frequency of measured AS values (averaged over 100 wavelengths). The median AS is approx. 5° for the high antenna location in Aarhus and 10–12° for the low antenna locations of Aarhus and Stockholm.

In Section 6, we will discuss how the AS impacts the spatial filtering gain of *conventional* beamforming.

4. Network Simulation Model and Signal Description

4.1. NETWORK SIMULATION MODEL

A simplified GSM network simulation model was developed in order to test and optimize various uplink and downlink beamforming algorithms under realistic network interference conditions [18]. The network model includes ten co-channel cells with a topology according to a 1/3 reuse network (tri-sectored base station configuration), see Figure 2. Cell #0 is the desired cell for the performance analysis and cells #1–9 are co-channel interfering cells. The path-loss model is $k \cdot r^{-3.5}$, where k is a constant and r the range. Log-normal fading with a standard deviation of 8 dB was applied. The network simulator can import both simulated and measured multipath channel data for the desired mobile in cell #0. Only simulated multipath data were used for the co-channel interfering signals from cells #1–9. A data file containing the time-space impulse responses along the trajectories sketched in Figure 2 was created for each cell. For the generation of time-space multipath data, the scatterer model described in [18] was used.
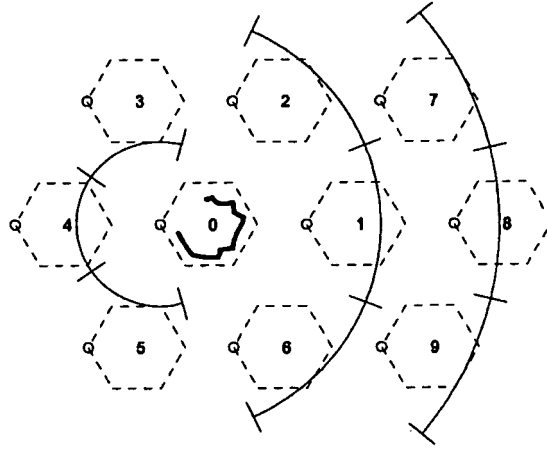
Figure 2. Cellular geometry of the co-channel model used for simulating a 1/3 reuse GSM network. The desired user follows a path in cell #0, while the interfering users in cells #1–9 move on the shown arcs.

4.2. GSM SYSTEM ASPECTS

The network simulator can be configured to both synchronous and asynchronous modes. In the synchronous mode, all cells are full-time synchronized and the differences in propagation delay are omitted. In the asynchronous mode, a uniform random timing offset is applied to each co-channel cell. The interfering cells employ other training sequences than the desired user. FH is also a feature in the network simulator. For the narrowband measurements, decorrelation in fading between successive bursts (frequency diversity) is emulated by performing short spatial steps around the true mobile location for each burst. For the wideband mode, the FH is emulated according to the description in Section 3.3. Uplink DTX is simulated, whereas uplink Power Control (PC) was disregarded in the analysis.

4.3. SIGNAL DESCRIPTION

The simulation model is based on the multidimensional impulse responses between every mobile in the network and the base station antenna array of the desired cell. The base station antenna topology is a uniform linear array; 8 elements with half a wavelength element separation. The co-channel interfering users move forward and backward on the arc segments shown in Figure 2. The desired user moves on a path in cell #0, for example as shown in Figure 2. The real measurement route determines the actual path. The received spatial data vector for symbol q in burst b is given by

$$\mathbf{x}_q = [x_{1,q}, x_{2,q}, \ldots, x_{i,q}, \ldots, x_{M,q}]^T. \tag{2}$$

The index b is omitted in Equation (2) and in the sequel for clarity. It is important to note that the signal sample vector $\mathbf{x}_q$ includes signal contributions from both the desired mobile and all 9 interfering mobiles.

5. Design of Beamforming Algorithm

This section describes the designed beamforming algorithm, which has been derived for the downlink case. We do, however, also suggest using the algorithm for uplink reception.

For downlink transmission, one would ideally like to know the instantaneous spatio-temporal impulse response matrix $\mathbf{H}^{TX}$ for every burst number b. This matrix contains the impulse response from the mobile antenna to each element in the base station antenna array. The dimension is $M \times N$, where N is the length of the temporal impulse response.

$$\mathbf{H}^{TX} = [\mathbf{h}_1^{TX}\mathbf{h}_2^{TX} \ \ldots \ \mathbf{h}_q^{TX} \ \ldots \ \mathbf{h}_N^{TX}]\,. \tag{3}$$

The vector $\mathbf{h}_q^{TX}$ is the transmit spatial impulse response vector for the excess delay index q in burst b. The excess delay index q in the sequel is also referred to as the *delay tap number*. The dimension of $\mathbf{h}_q^{TX}$ is $M \times 1$. Unfortunately, with respect to obtaining $\mathbf{H}^{TX}$, GSM uses frequency division duplexing and therefore the instantaneous value of $\mathbf{H}^{TX}$ cannot be found from the corresponding uplink estimate $\mathbf{H}^{RX}$. Nevertheless, under the assumption of uncorrelated scattering, the expectation value $E\{|\mathbf{H}^{TX}|\}$ is identical to $E\{|\mathbf{H}^{RX}|\}$.

Since we are using a random FH-GSM network, it is difficult for the base station beamformer to achieve directional information about co-channel mobile stations. This is because the downlink TDMA frame structure is 3 burst periods ahead of the corresponding uplink TDMA frame structure [31]. Downlink null steering techniques for the suppression of emitted interference to co-channel mobiles in adjacent cells [21] has therefore not been investigated. With the lack of positioning information for co-channel mobiles, it is desirable to use the transmit weight vector $\mathbf{w}$ maximizing the energy received at the desired mobile station over the total transmitted energy. The following performance criterion was used for computing the transmit vector $\mathbf{w}$ (index 0 corresponds to the desired mobile in cell #0 of the simulation model):

$$\mathbf{w} = \arg_{\mathbf{w}} \max \left(E \left\{ \frac{\mathbf{w}^H \mathbf{R}^{TX,0} \mathbf{w}}{\mathbf{w}^H (\mathbf{J}^H \cdot \mathbf{J}) \mathbf{w}} \right\} \right)\,, \tag{4}$$

where $\mathbf{R}^{TX,0} = \sum_q (\mathbf{h}_q^{TX,0})^H \cdot \mathbf{h}_q^{TX,0}$ is the sum of multipath covariances of the desired mobile and $\mathbf{J}$ the identity matrix. Here, H means transpose and complex conjugated. In [18], it was argued that for most radio channel conditions, less than 1 dB is lost by restricting the search of the transmit vector to the form

$$\mathbf{w} = \text{const} \times \mathbf{a}(\theta_0, f^{TX})\,, \tag{5}$$

which implies that the weighting vector is simply a scaling of a steering vector $\mathbf{a}(\theta, f)$ with the azimuth direction θ_0 and that f^{TX} is the downlink frequency. The steering vector $\mathbf{a}(\theta, f)$ is given by

$$\mathbf{a}(\theta, f) = [1, \exp(-j2\pi f \Delta \sin(\theta)/c), \ \ldots, \ \exp(-j2\pi f(M-1)\Delta \sin(\theta)/c)]^T\,, \tag{6}$$

where Δ is the antenna element spacing and c the speed of light. The constraint on $\mathbf{w}$ given in Equation (5) reduces the degrees of freedom from $2(M-1)$ to only one: θ. The weight vector of the form given in Equation (5) is also well known as *conventional* beamforming [22]. As no downlink entities can be estimated at the base station to solve Equation (4), we simply replace *TX* by *RX*. This approximation will not lead to any substantial performance degradation if the beam pattern for up- and downlink are similar. The best transmit direction is given by:

$$\theta_0 = \arg_\theta \max \left(E \left\{ \sum_q |\mathbf{a}(\theta, f^{RX})^H \mathbf{h}_q^{RX,0}|^2 \right\} \right)\,. \tag{7}$$

Equation (7) can also be interpreted as a *low resolution* Direction of Arrival (DoA) estimator. The more practical implementation of this DoA algorithm for GSM is described in the following section.

## 5.1. DoA Estimation Algorithm for GSM

This section describes the practical implementation of the DoA algorithm for GSM including the impulse response estimate $\hat{\mathbf{h}}_q^{RX,0}$ and the expectation $E\{\}$. The GSM traffic burst contains a midamble of 26 known symbols, the training sequence. The impulse response estimate $\hat{\mathbf{h}}_q^{RX,0}$ is obtained by a cross-correlation between the received signal and a Minimum Shift Keying (MSK) mapped version of the known training sequence $I^{RX,0}$ (16 or 26 bits may be used).[2]

$$\hat{\mathbf{h}}_s^{RX} = \sum_{q=s}^{s+25} \frac{1}{26} \mathbf{x}_q \left(I_{q-s}^{RX,0} \right)^*, \tag{8}$$

where $\mathbf{x}_q$ is the spatial signal vector given in Equation (2) and $*$ the complex conjugation. For each antenna element, the estimated impulse response is $N = 11$ symbols long (± 5 symbols time lag) and the dimension of $\hat{\mathbf{H}}_s^{RX,0}$ becomes $M \times 11$. A crucial issue for the impulse response estimation is the poor cross-correlation properties of the GSM training sequences. This is especially an issue for a synchronized network. For equalization in GSM, it is typically assumed that the Power Delay Profile is within 5 symbols. However, in order to reduce the impact of co-channel interference on the DoA estimate (due to poor cross-correlation properties), it is here further dictated that most of the energy in the impulse response received from the direction θ_0 is kept within a window of only 3 symbols $s \in \{k_0, k_0 + 1, k_0 + 2\}$ where $k_0 = 1, \ldots, 9$. This constraint is based on the fact that the pulse-shaping filter for the linear MSK/OQAM model of GMSK has an *effective* duration of approx. 2 symbol periods and the assumption that most energy received from a certain azimuth direction is within 1 symbol duration (3.7 μs). This assumption is valid for both the TU scenario of Aarhus and the BU two-cluster scenario of Stockholm. Here, the energy received from the second cluster with a long excess delay has a different azimuth direction than the energy received from the 1st cluster with low excess delay. The expectation $E\{\}$ in Equation (7) is implemented by averaging over B bursts. From analysis of the burst pattern in DTX mode of GSM, it was suggested in [18] to use $B = 21$. This corresponds to an averaging window of approx. 100 ms in the non-DTX mode and approx. 520 ms (slightly more than a SACCH multiframe [31]) in the DTX mode. It should be noted that for slow-moving or stationary mobile stations, the time averaging over B bursts does not ensure any averaging over fast fading unless FH is used to obtain frequency diversity. Hence, FH is very essential for the performance of downlink beamforming in a GSM network with a mixture of fast- and slow-moving mobile stations.

When using random FH, the uplink log-normal fading process in a combination with uplink DTX and PC of co-channel interfering mobiles will accommodate a very large variation in interference level. Some bursts will be received as heavily interfered, whereas others will be received almost free of interference. Because of the non-ideal cross-correlation properties of the GSM training sequences, a "hit" by a strong co-channel interfering signal may sometimes result in an erroneous estimation of θ_0 even when averaging over B bursts [18]. This has been the motivation for applying logarithmic power averaging instead of linear, as the latter has

---

[2] 26 bits give better noise and interference suppression than 16 bits, but the autocorrelation properties worsen.
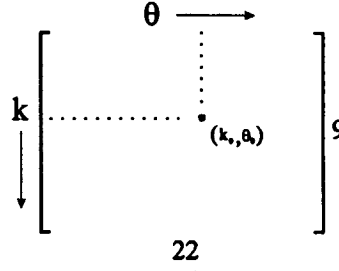
Figure 3. Selection of fixed beam direction θ_0.

been found more resistant to this sort of interference distribution, see Equation (9). For the considered case of an 8-element ULA covering a sector of 120°, [18] decided to use a grid of beams with a crossover depth of 0.5 dB between adjacent beams in order to reduce the computational load of the algorithm. This constraint resulted in 22 fixed beam directions for θ_0

$$\Theta = [-72.7, \ -59.4, \ -49.8, \ -41.9, \ -35.0, \ -28.5, \ -22.6, \ -16.6, \ -11.0,$$
$$-5, 5, \ 0.0, \ 5.5, \ 11.0, \ 16.6, \ 22.6, \ 28.5, \ 35.0, \ 41.9, \ 49.9, \ 59.4, \ 72.7, \ 88.2].$$

The time averaged received power $P(k, \theta)$, $k \in \{1, \ldots, 9\}$ and $\theta \in \Theta$ are now given by:

$$P(\theta, k) = \frac{1}{B} \sum^{B} \log \left(\sum_{s=k}^{k+2} \left| a^H (\theta, f^{RX}) \hat{h}_s^{RX,0} \right|^2 \right),$$ (9)

while the azimuth direction θ_0, which is selected for the beamforming transmit vector $a(\theta_0, f^{TX})$, is given by

$$\theta_0 = \arg_{\theta \in \Theta, k \in [1, \ldots, 9]} \max P(\theta, k).$$ (10)

The selection process of Equation (10) is illustrated in Figure 3. There are 9 possible values of k (excess delay) and 22 fixed values of θ (azimuth direction).

The computational and memory requirements of the algorithm are rather low relative to the performance of modern DSPs. We do, therefore, believe that the algorithm is feasible for real-time implementation.

5.2. ASPECTS OF ANALOGUE BEAMFORMING

In the derivation of the beamforming algorithm presented in the previous section, we assumed a digital beamforming implementation (which is the case for the TSUNAM II adaptive antenna base station). However, the concept of digital transmit beamforming may not be cost efficient for commercial implementation. Analogue beamforming, by e.g., using a Butler matrix, is an obvious alternative [19]. Fortunately, because of the constraints on the transmit vector w set in Equation (5), the algorithm is also applicable for an analogue beam-switching approach. In Section 6.3, we also include results for both an 8 × 8 and a 4 × 4 Butler matrix implementation. The steering directions for the 8 × 8 Butler matrix are:

$$\Theta = [-60.7, \ -38.7, \ -22.5, \ -7.2, \ 7.2, \ 22.5, \ 38.7, \ 60.7].$$ (11)

For a 4 × 4 Butler matrix, the following fixed beam directions are:

$$\Theta = [-48.6, \ -14.5, \ 14.5, \ 48.6].$$ (12)

The crossover depth between adjacent beams for both the 4 × 4 and 8 × 8 Butler matrix solutions is approx. −4 dB, which may significantly degrade the average performance of the adaptive antenna. For the 8 × 8 implementation, the crossover depth is narrow compared to the azimuth spread measured in urban environments and the depth will partly be filled by multipath signals. This is different for a 4 × 4 Butler matrix implementation because of the wider depths and we therefore suggest the use of two Butler matrix implementations with a half beamwidth offset in beam directions, resulting in 8 beam directions. Such an implementation can be found in [20] for the case of two 6 × 6 Butler matrixes using orthogonal polarization.

6. Uplink Simulation Results

Although the beamforming algorithm described in the previous section was derived for downlink, we will first look into its uplink performance. This is partly because the DoA estimation is an uplink issue and, secondly, because the layout of the simulated network is better suited for investigating uplink performance. Downlink performance results will be given in Section 7.

6.1. DoA Characteristics

The designed DoA algorithm should preferably estimate the azimuth direction of the strongest signal path from the desired user (averaged over short-term fading). For low AS values, the *mean* DoA estimate is mostly identical to the geometrical azimuth of the mobile station with some variation. However, in environments with large azimuth spread (such as the Bad Urban, two-cluster scenario from Stockholm), the DoA estimate can be significantly biased from the geometrical azimuth of the mobile station. Figure 4 gives an example comparing the geometrical and estimated DoA as a function of time. The measurement route is from Aarhus, repeated for the 3 different base heights: 32 m, 26 m, and 20 m (note the vertical shift of the curves). It can be observed that the mean of the DoA estimate follows the geometrical azimuth for all three heights, but the variation in DoA estimates increases as the antenna height is lowered (corresponding to an increase in AS).

6.2. Beamforming Gain Dependence on Single Element C/I

The DoA algorithm was designed with the goal of being robust to interference. If simply applying $10 * \log_{10}(M)$ for the maximum beamforming gain where $M = 8$ and assuming a required C/I of 9 dB at the input of the detector [31], then the worst case single element C/I should at least be 0 dB. This worst case C/I of 0 dB is an average over multipath fading, and hence approximately half the bursts will be received with a negative instantaneous single element C/I value. Figures 5(a,b) show the DoA estimates for a small part of the Stockholm route for the cases of −2 dB and −12 dB C/I, respectively.

Figure 5 clearly shows an enlarged fluctuation in the DoA estimate when the C/I is lowered from −2 dB to −12 dB. For very low C/I, the DoA algorithm sometimes tracks strong co-channel interfering mobiles instead of the desired mobile. In Figure 6, the beamforming gain (over single element) is shown as a function of C/I (for the same part of the route as shown in Figure 5). It can be observed that the gain significantly decreases when the C/I becomes lower than approx. −8 dB. This is, however, not fatal, as the beamformer will operate at single element C/I of 0 dB or more as discussed above. The length of the GSM training sequence (i.e. cross-correlation properties) and the length of the burst averaging B mainly determine the "breakdown" point of the algorithm.
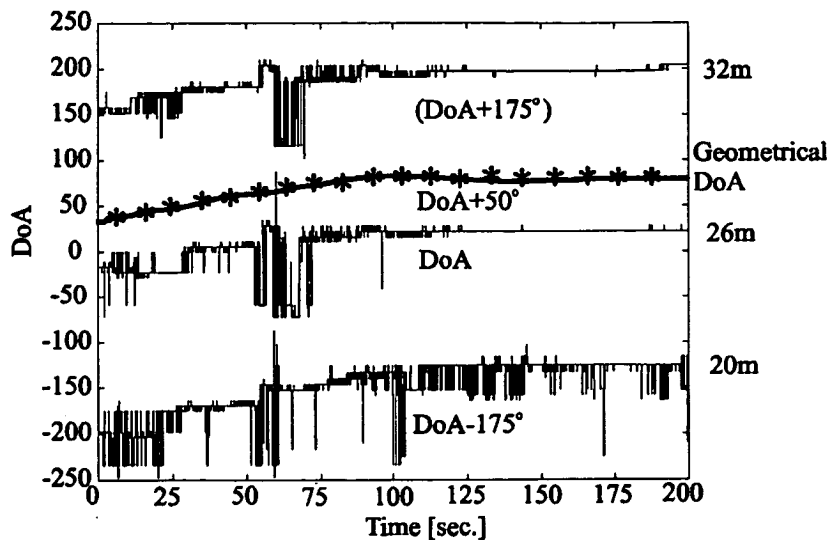
Figure 4. Characteristics of the DoA for the Aarhus measurement route at 3 different base station antenna heights.
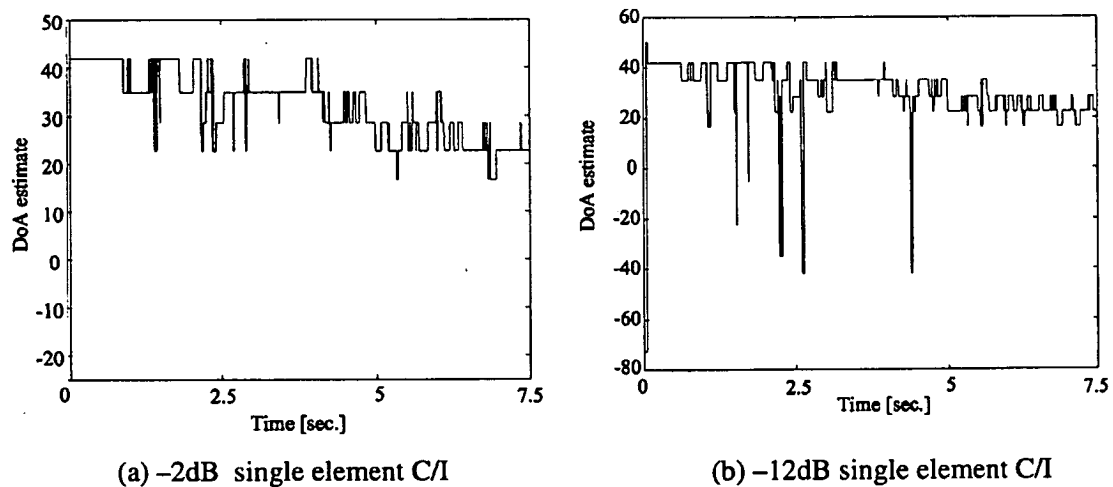

(a) –2dB single element C/I   (b) –12dB single element C/I

Figure 5. DoA estimate for a small part of the Stockholm measurement route: (a) C/I = –2 dB; (b) C/I = –12 dB.

6.3. C/I GAIN DISTRIBUTION

Figure 7 shows the simulated C/I improvement for the Aarhus measurement route for the antenna height of 32 m. As mentioned, the azimuth sweep of the measurement route ranges from +20° to –30° and the distance to the base station varies between 1.5 km and 2 km. For clarity, the C/I improvement is shown as an average over 40 bursts.

It can be observed from Figure 7 that the local C/I improvement varies significantly along the measurement route, i.e., the improvement is position-dependent. It can also be observed that the C/I improvement is almost uncorrelated with the single element C/I, which is usually better than –5 dB (note the vertical shift of the curve by –10 dB in the plot!). Figure 8 shows the cumulative frequency of C/I values, averaged over 8 bursts, for the Aarhus measurement route
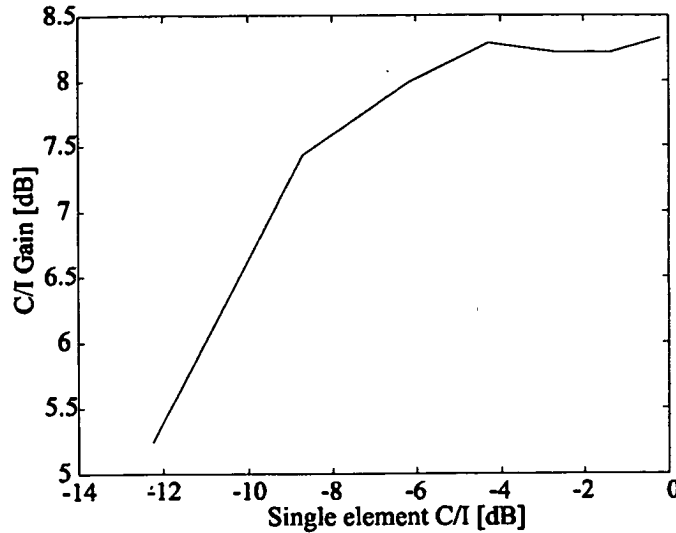
Figure 6. Mean beamforming gain (over single element) as a function of C/I.
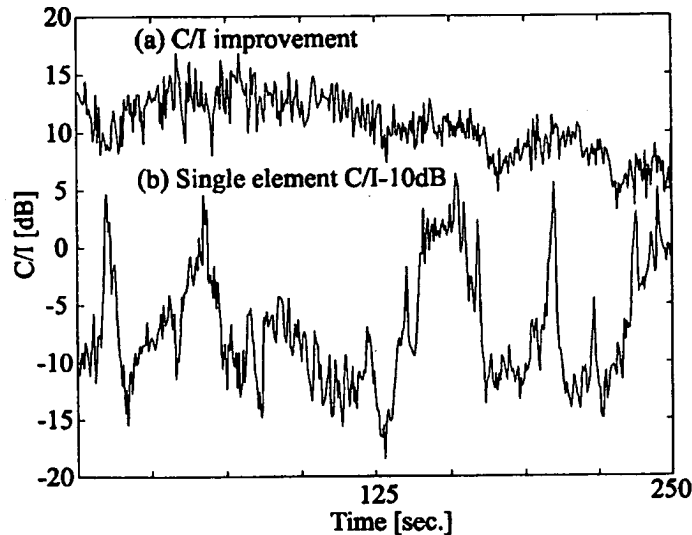
Figure 7. (a) C/I improvement (over single element) versus time for the antenna height of 32 m, Aarhus. (b) Single element C/I.

for both single element and for beamforming ($M = 8$). The 8-burst averaging period corresponds to the interleaving period of the GSM full rate speech channel. We chose to use the 10% outage level to identify the beamforming gain. Hence, from Figure 8, the beamforming gain is 8.3 dB.

Table 1 summarizes the C/I gain of different antenna configurations for an outage level of 10%. The results are shown for both measured channel data from Aarhus and for simulated radio channel data (having the same median AS). The results from Aarhus show that a 4-element antenna array is sufficient to obtain a beamforming gain on the order of 5 dB, provided the number of fixed steering directions is 8 (or more), hence facilitating a full load in a 1/3 reuse FH-GSM network (see Section 2). For the case of AS = 5°, there is a very good agreement

Figure 8. Cumulative frequency of the single element C/I and beamforming C/I ($M = 8$) for the Aarhus measurement route and 32 m base station height.

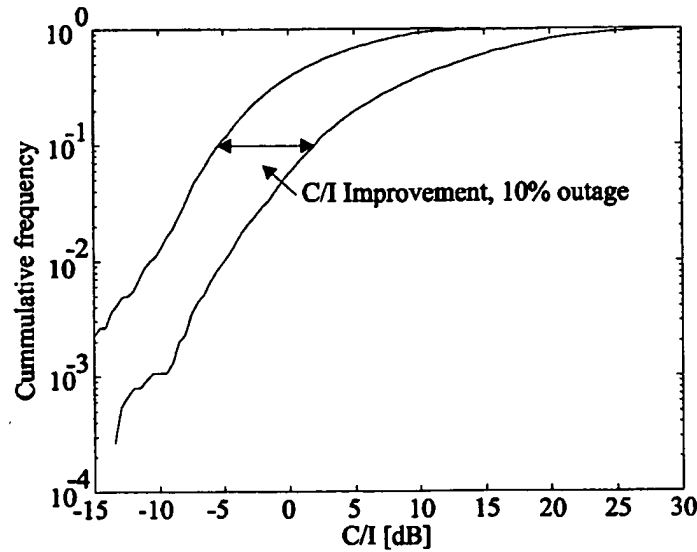*Table 1.* The C/I gain at the 10% outage level for different antenna array configurations: AS = median azimuth spread; F = number of fixed beam directions. The measured results are from Aarhus. The simulated results were generated using the TU model in [18].

Base height	32 m, (AS = 5°)		20 m, (AS = 10°)	
	Relative C/I (dB), simulated	Relative C/I (dB), measured	Relative C/I (dB), simulated	Relative C/I (dB), measured
Single elem.	0	0	0	0
$M = 8, F = 22$	8.7	8.3	5.9	7.4
$M = 8, F = 8$	7.6	7.8	5.8	6.6
$M = 4, F = 8$	5.5	5.8	4.6	5.3
$M = 4, F = 4$	5.2	4.8	3.4	4.5

between the measured and simulated radio channel data, whereas for the case of AS = 10° the discrepancy is on the order of 1–1.5 dB. It should be noted that the simulated TU channel data assumes a Gaussian PAS [18] instead of a Laplacian one, as later found in [15]. The Gaussian model seems to predict too pessimistic results for conventional beamforming.

Figure 9 shows similar results as Figure 8, but for the severe BU Stockholm route. For a part of this route, two separate scattering clusters located at –30° and +25° in azimuth with almost equal mean power were observed [27]. At the 10% outage level, the beamforming gain is only 5.3 dB, which is a significant reduction compared to the results for the Aarhus route. Furthermore, the curve for beamforming starts to approach the single element curve for a cumulative frequency of less than 5%. These results from Stockholm indicate that it
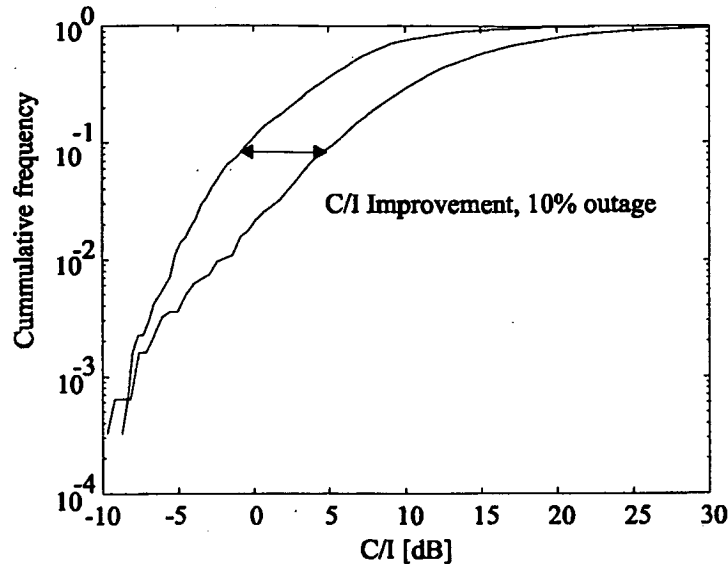
Figure 9. Cumulative frequency of the single element C/I and the beamforming C/I ($M = 8$) for the Stockholm measurement route.

can become difficult to obtain large gain values of *conventional* beamforming in some severe urban propagation environments.

6.4. ANALYTICAL RESULTS FOR CONVENTIONAL BEAMFORMING

When the AS becomes large relative to the antenna beamwidth of conventional beamforming, all of the impinging waves will not be captured and the Array Factor (AF) becomes saturated. In Figure 10, the relative loss in captured signal power is shown as a function of the 3 dB beamwidths for different AS. The curves are computed analytically, assuming a Laplacian PAS. The *ideal* AF for a single wave is also shown in the figure (dashed line).

For example: for $M = 8$, the *ideal* AF is 9 dB and the 3 dB beamwidth is approximately 13°. For an AS of 10°, the captured power is reduced by 2.2 dB. Hence, the *effective* AF is reduced to (9–2.2) dB = 6.8 dB. Similarly for $M = 4$ and AS = 10°, the *effective* AF is (6–0.8) dB =5.2 dB. The additional gain from doubling the number of antenna elements from 4 to 8 is only 1.6 dB. In urban areas, therefore, we suggest the limiting of the number of antenna elements to 6–8 for cost-efficiency reasons. Inspection of Figure 10 for $M = 8$ and $M = 4$ gives the following *effective* AF values at 5° and 10° azimuth spread:

- $M = 8, \sigma = 5°$: (9–0.8) dB = 8.2 dB
- $M = 8, \sigma = 10°$: (9–2.2) dB = 6.8 dB
- $M = 4, \sigma = 5°$: (6–0.2) dB = 5.8 dB
- $M = 4, \sigma = 10°$: (6–0.8) dB = 5.2 dB

By comparing these gain values, obtained by simple analytical calculations, to the values obtained by extensive network simulations (see Table 1), a very good agreement can be found. Hence, when the median AS of a certain environment is known, the performance of adaptive base station antennas using *conventional* beamforming can be accurately predicted using Figure 10 and thereby avoiding heavy computer simulations and time-consuming test
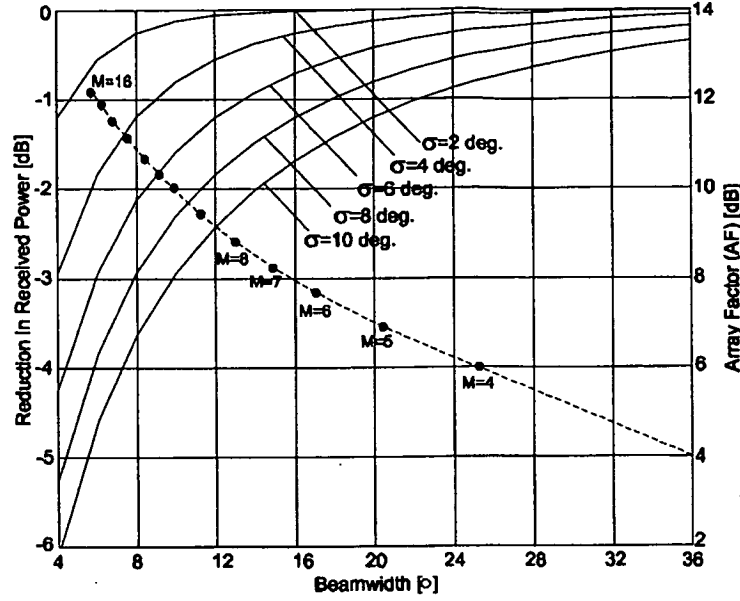
Figure 10. Array factor and relative signal power loss versus antenna beamwidth for different values of azimuth spread.

measurements. In [28, Figure 6], it was suggested to predict the AS from cross-correlation statistics of (existing) 2-branch space diversity configurations.

7. Downlink Simulation Results

In this section, the performance improvement for downlink is analyzed. For the downlink performance analysis, only simulated channel data have been used. This was in order to ensure a uniform azimuth location of the desired user in cell #0. The following performance measure was used:

$$G_{L,i} = \frac{C_{L,L}^{Array} \cdot I_{L,i}^{Single}}{I_{L,i}^{Array} \cdot C_{L,L}^{Single}},$$ (13)

where $C_{L,L}$ is the power delivered by the serving base station in cell L to the desired mobile in cell L and $U_{L,i}$ is the interfering power delivered from the serving base station in cell L to a co-channel mobile in cell i. In the simulator, L and i are restricted to 0 and 1, respectively. The superscripts "Single" and "Array" denote single element and antenna array cases, respectively. All values in Equation (13) have been averaged over fast fading (40 bursts). Calculation of the instantaneous powers is done as follows:

$$C = |w^H \cdot H_{L,L}^{TX}|^2$$ (14)

$$I = |w^H \cdot H_{L,i}^{TX}|^2,$$ (15)

where $H_{L,L}^{TX}$ is the impulse response matrix between the serving base station and the desired mobile in cell L and $H_{L,i}^{TX}$ is the impulse response matrix for the serving base station in cell L to the co-channel mobile in cell i. w is given by Equation (5). The downlink interference
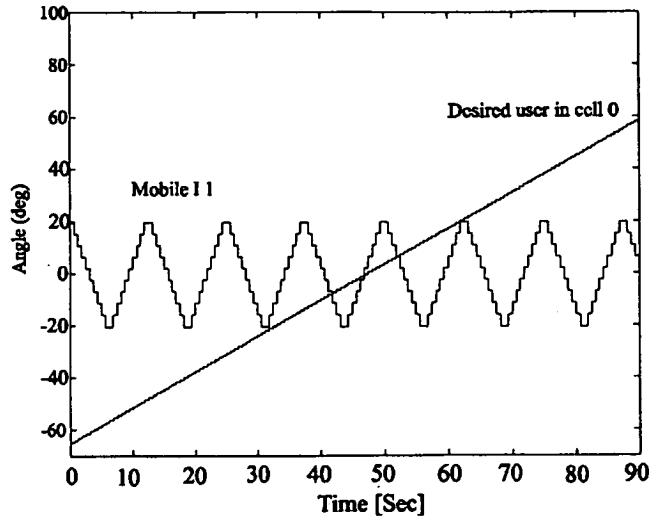
Figure 11. Azimuth of the desired mobile in cell #0 and the interfered mobile "I1" in cell #1 versus time.
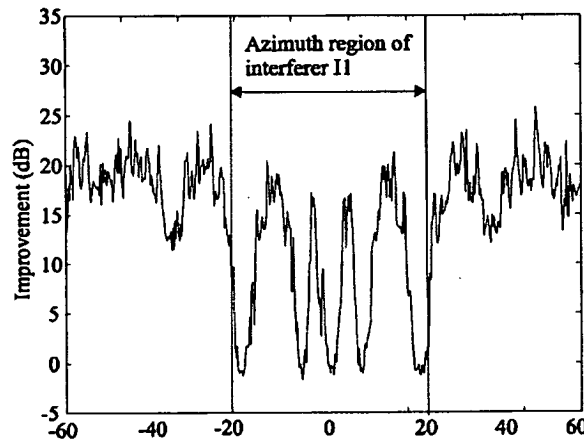
Figure 12. C/I improvement according to Equation (13). The azimuth region in which "I1" potentially gets interfered is shown by the vertical lines.

improvement for the interfered mobile in cell #1 ($i = 1$) was analyzed. The azimuth of the desired mobile in cell #0 and the interfered mobile "I1" in cell #1 are shown in Figure 11. The desired mobile performs an azimuth sweep of –60° to 60°, while the interfered mobile moves between –20° and 20° relative to the base station in cell #0. It can be observed that the interfered mobile in cell #1 has a larger angular speed than the desired mobile in cell #0, which of course is unrealistic. The reasoning for this is to ensure better statistical output from a short Monte Carlo simulation, and it has no significant influence on the mean performance results.

From Figure 12, it can be observed that when the desired mobile in cell #0 is in the same azimuth region as the interfered mobile in cell #1, the performance improvement of *conventional* beamforming is almost an on/off function. This is expected, because when the azimuth separation between the interfered and desired mobile is less than half the beamwidth (AS = 0°), the beamformer provides no or only little interference suppression. For azimuth

Table 2. Downlink performance improvement (over single element) for different configurations of the base station antenna array (AS = 7.5°).

Configuration	4 Antennas		8 Antennas	
	4 Beams	8 Beams	8 Beams	22 Beams
C/I gain	4.1 dB	5.3 dB	7.3 dB	7.8 dB

separations of more than half the beamwidth, the interference suppression is determined by the side lope level. If a non-uniform window function is used, the side lope level may be improved for the penalty of increased beamwidth. For a random FH network, our simulation results have shown that a uniform window function gives the best mean performance.

The interference situation shown in Figure 12 is without FH. In a random FH network, the interferer situation will change from time slot to time slot and effectively perform an averaging of the interference improvement pattern. The average improvement over a single element is 7.8 dB for the simulation of digital beamforming with 22 fixed beams and $M = 8$ (AS = 7.5°).

Improvements for the different beamforming and antenna array configurations are summarized in Table 2. The simulated downlink results shown in Table 2 are in good agreement with the uplink results shown in Table 1.

7.1. FIELD TRIAL RESULTS

The proposed beamforming algorithm has been implemented and tested *live* in the TSUNAMI II, GSM-1800 adaptive antenna base station demonstrator at Orange PCS Ltd in Bristol, U.K. An extensive description of the field trial and the results can be found in [24] and [25]. It was concluded in [24] that the described beamforming algorithm (referred to in [24] as "(TRB) Grid of Beams"), among other algorithms, showed the best overall link performance in terms of quality. In the micro-cellular field trial, the mean downlink improvement was estimated at 5.7 dB ($M = 8$). This result is in the same range as our results from Stockholm. Considering the relative large AS expected for a micro-cellular base station installation, this is a satisfactory result for the algorithm.

8. Conclusion

A robust DoA algorithm for downlink *conventional* beamforming in random FH-GSM has been presented. The algorithm was thoroughly verified by means of network computer simulations using both simulated and measured radio channel data. It was also tested in a GSM-1800 field trial in the U.K. with good results. The Azimuth Spread (AS) was found to be an important parameter for the performance of *conventional* beamforming. The simple $10 * \log 10(M)$ expression for the spatial filtering gain of a point source was too optimistic for urban areas. Space-time measurements were conducted in the two dissimilar cities of Aarhus and Stockholm, where the median azimuth spread was in the range of 10°–12° for low macro-cellular base station antenna heights (approx. 20 m). The gain for an 8-element array was reduced from the theoretical value of 9 dB for AS = 0° to 7.4 dB and 5.3 dB for the Aarhus and Stockholm

measurement results, respectively. We conclude that 4 to 6 antenna elements provide a spatial filtering gain on the order of 5 dB in most urban macro-cellular environments, thus allowing a 1/3 reuse FH-GSM network to be fully loaded and thereby achieve a capacity increase of $\times 3$.

Acknowledgements

References

1. S. Anderson, U. Forssén and J. Karlsson, "Ericsson/Mannesmann GSM Field-Trials with Adaptive Antennas", *Proc. of EPMCC'97*, pp. 77–85.
2. M. Tangeman, U. Bigalk, C. Hoek and M. Hother, "Sensitivity Enhancements of GSM/DCS 1800 with Smart Antennas", *Proc. of EPMCC'97*, pp. 87–94.
3. A.J. Poulraj and C.B. Papadias, "Space-Time Processing for Wireless Communications", *IEEE Signal Processing Magazine*, pp. 49–83, 1997.
4. J.H. Winters, "Smart Antennas for Wireless Systems", *IEEE Personal Communications*, pp. 23–27, 1998.
5. R. Kohno, "Spatial and Temporal Communication Theory Using Adaptive Antenna Array", *IEEE Personal Communications*, pp. 28–35, 1998.
6. F. Kronestedt and S. Andersson, "Migration of Adaptive Antennas into Existing Networks", *Proc. of IEEE VTC'98*, Ottawa, pp. 1670–1674.
7. U. Forssén, J. Karlsson, B. Johannisson, M. Almgren, F. Lotse and F. Kronestedt, "Adaptive Antenna Arrays for GSM900/DCS1800", *IEEE Proc. of VTC '94*, Stockholm, Sweden, pp. 605–609.
8. P. Kenington and W. Slingsby, "Multi-Carrier Linear Power Amplifier Design for a DCS1800 Adaptive Antenna System", *Proc. of TSUNAMI II Final Workshop on Adaptive Antennas*, Copenhagen, Sept. 1998.
9. F. Kronestedt and S. Andersson, "Migration of Adaptive Antennas into Existing Networks", *IEEE Proc. of VTC'98*, Ottawa, May 1998, pp. 1670–1674.
10. Technical Annex, "Technology in Smart Antennas for Universal Advanced Mobile Infrastructure Part II (TSUNAMI 2)", EU, ACTS Project Number AC020, March 1998.
11. "Traffic Loading of GSM Networks", European Radiocomunications Office, April 1998.
12. W.C. Lee, *Mobile Communications Engineering*, McGraw-Hill, Inc., 1976.
13. F. Kronestedt and S. Andersson, "Adaptive Antennas in Frequency Hopping GSM", *IEEE Proc. of ICUPC*, Oct. 1998, pp. 325–329.
14. F. Frederiksen, P.E. Mogensen, K.I. Pedersen and P.L. Espensen, "Testbed for Evaluation of Adaptive Antennas for GSM and Wideband CDMA", *Proc. of COST 252/259 Joint Workshop*, Bradford, April 21–22, 1998, pp. 171–181.
15. K.I. Pedersen, P.E. Mogensen and B.H. Fleury, "Power Azimuth Spectrum in Outdoor Environments", *IEE Electronics Letters*, Vol. 33, No. 18, pp. 1583–1584, 1997.
16. U. Martin, "A Directional Radio Channel Model for Densely Built-Up Urban Areas", *Proc. 2nd EPMCC*, Bonn, Germany, Oct. 1997, pp. 237–244.
17. M. Tangemann and M. Beach et al., "Report on the Benefits of Adaptive Antennas for Cell Architectures", RACE TSUNAMI R2108/SEL/WP3-4/DS/P/029/b1, May 1995.
18. P.E. Mogensen, P. Zetterberg, H. Dam, P.L. Espensen and F. Frederiksen, "Algorithms and Antenna Array Recommendations", Technical Report AC020/AUC/A1.2/DR/P/005/b1, 1997.
19. J.P. Shelton and K.S. Kelleher, "Multiple Beams from Linear Arrays", *IRE Trans. on Antennas and Propagation*, pp. 154–161, 1961.
20. B. Johannisson, B., "Active and Adaptive Base Station Antennas for Mobile Communication", *Proc. of Antenna'97*, Sweden.
21. P. Zetterberg, "Mobile Cellular Communications with Base Station Antenna Arrays: Spectrum Efficiency, Algorithms and Propagation Models", Ph.D. Thesis, KTH, Stockholm, ISSN 1103-8039, 1997.

22. B.D. Veen and K.M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering", *IEEE ASSP Magazine*, pp. 4–24, 1988.
23. J.H. Winters, "Optimum Combining in Digital Mobile Radio with Co-Channel Interference", *IEEE Journal on Selected Areas in Communications*, Vol. SAC-2, No. 4, 1984.
24. R. Arnott and S. Ponnekanti, "Experimental Investigation of Adaptive Antennas in a DCS-1800 Mobile Network", *Proc. of 3rd ACTS Mobile Communication Summit*, Rhodes, Greece, June 8–11, 1998, pp. 535–543.
25. T. Howard, C.M. Simmonds, P. Darwood, M.A. Beach, R. Arnott, F. Cesbron and M. Newman, "Adaptive Antenna Performance in Mobile Systems", Report AC020/ORA/WP3/DS/P/008/b1, June 30, 1998.
26. M. Pokkila and P. Ranta, "Channel Estimator for Multiple Co-Channel Demodulation in TDMA Mobile Systems", *Proc. of EMPCC'97*, Bonn, Sept. 1997.
27. K.I. Pedersen, P.E. Mogensen and B.H. Fleury, "A Stochastic Model of the Temporal and Azimuthal Dispersion Seen at the Base Station in Outdoor Propagation Environments", *IEEE Trans. on Vehicular Technology* (submitted), 1998.
28. K.I. Pedersen, P.E. Mogensen and B.H. Fleury, "Spatial Channel Characteristics in Outdoor Environments and their Impact on BS Antenna System Performance", *IEEE Proc. of VTC'98*, Ottawa, Canada, May 18–21, 1998, pp. 719–723.
29. S. Ratnavel, A. Paulraj and A.G. Constantinides, "MMSE Space-Time Equalization for GSM Cellular Systems", *IEEE Proc. of VTC '96*, Atlanta, April 1996, pp. 331–335.
30. J.H. Winters, "The Impact of Antenna Diversity on the Capacity of Wireless Communications Systems", *IEEE Trans. on Communications*, Vol. 42, No. 2/3/4, pp. 1740–1751, 1994.
31. European Digital Cellular Telecommunications System, ETSI, GSM 05.XX, ETS 300 XXX.
32. UTRA UMTS Physical Layer Description (v04, 1998-06-25), 1998.
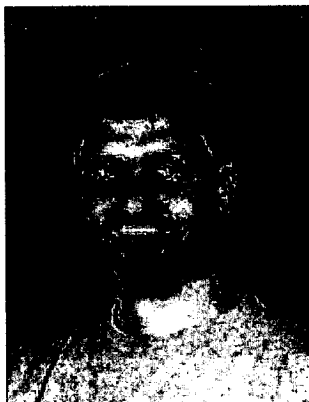
Preben E. Mogensen received his M.Sc. in electrical engineering and Ph.D. degrees in 1988 and 1996, from Aalborg University, Denmark. Since 1988, he has employed at Aalborg University, where he is currently an associated research professor at the research centre, *Center for Personkommunication* (CPK). He is leading the Cellular Systems research group in CPK. Since 1995 he has been part time System Consultant for Nokia Telecommunications. His current research areas are adaptive antenna technology and radio resource management for 2nd and 3rd generation cellular systems.

Klaus Ingemann Pedersen received his M.Sc. degree in electrical engineering in 1996 from Aalborg University, Denmark. His M.Sc. thesis was completed at Wireless Information Network Laboratory (WINLAB), Rutgers University, NJ. U.S.A. During 1996 he worked at Lucent Technologies Bell labs innovations where he was involved in design of CDMA receivers. He is currently a Ph.D. student at Center for PersonKommunikation, Aalborg University. His current research interests are 2-dimensional channel modeling and adaptive array signal processing for wideband CDMA systems.

Frank Frederiksen received his M.Sc. degree in electrical engineering in 1994, from Aalborg University (AAU), Denmark. He has since then been employed at Center for Personkommunikation, at AAU. His current working area is testbed development for adaptive antenna array field trials for wideband CDMA systems.

Poul Leth-Espensen received his M.Sc. degree in electrical engineering in 1994 from Aalborg University (AAU), Denmark. From 1994 to 1998 he was employed at Center for Personkommunikation at AAU. His research area was algorithm development for microphone and antenna arrays. In 1998 he joined Terma Elektronik A/S, Denmark where he is currently working on radar receiver development.
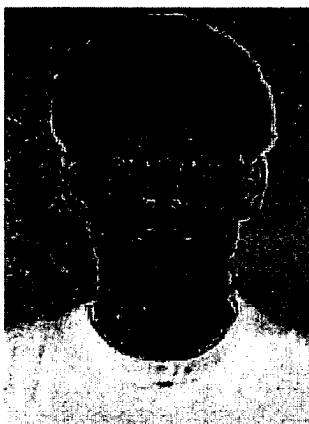
Per Zetterberg was born in Uppsala, Sweden, 1968. He received the M.Sc. degree in electrical engineering from Luleå University, Sweden, in 1993. In 1997 he received the Ph.D. degree in electrical engineering from the Royal Institute of Technology (KTH). During 1995–1996 he was a visiting researcher at Aalborg University, Denmark. Since June 1997, he has been with Radio Design AB as a research engineer. Zetterberg's research interests are in the areas of wireless communications and antenna array signal processing.

Data Gathering in Sensor Networks using the *Energy*Delay* Metric

Stephanie Lindsey*+, Cauligi Raghavendra*, and Krishna Sivalingam+

*Computer Systems Research Department
The Aerospace Corporation
P. O. Box 92957
Los Angeles, CA 90009-2957

+School of Electrical Engineering and Computer Science
Washington State University
Pullman, WA 99164-2752

Abstract

*In this paper we consider the problem of data collection from a sensor web consisting of N nodes, where nodes have packets of data in each round of communication that need to be gathered and fused with other nodes' packets into one packet and transmitted to a distant base station. Nodes have power control in their wireless communications and can transmit directly to any node in the network or to the base station. With unit delay cost for each packet transmission, if all nodes transmit data directly to the base station, then both high energy and high delay per round will occur. In our prior work [6], we developed an algorithm to minimize the energy cost per round, where a linear chain of all the nodes are formed to gather data, and nodes took turns to transmit to the base station. If the goal is to minimize the delay cost, then a binary combining scheme can be used to accomplish this task in about log N units of delay with parallel communications and incurring a slight increase in energy cost. The goal is to find data gathering schemes that balance the energy and delay cost, as measured by energy*delay. We conducted extensive simulation experiments with a number of schemes for this problem with 100 nodes in playing fields of 50m x 50m and 100m x 100m and the base station located at least 100 meters and 200 meters, respectively, from any node. With CDMA capable sensor nodes, a chain-based binary scheme performs best in terms of energy*delay. If the sensor nodes are not CDMA capable, then parallel communications are possible only among spatially separated nodes, and a chain-based 3 level hierarchy scheme performs well. These schemes perform 60 to 100 times better than direct scheme and also outperform a cluster based scheme, called LEACH [3].*

1. Introduction

Inexpensive sensors are deployed for data collection from the field in a variety of scenarios including military surveillance, building security, in harsh physical environments, for scientific investigations on other planets, etc. [2,4,13]. A sensor node will have limited computing capability and memory, and it will operate with limited battery power. These sensor nodes can self organize to form a network and can communicate with each other in a wireless manner. Each node has transmit power control and an omni-directional antenna, and therefore can adjust the area of coverage with its wireless transmission. For example, a sensor network can be used for detecting the presence of potential threats in a military conflict. Since wireless communications consume significant amounts of battery power, sensor nodes should be energy efficient in transmitting data [5,10,12]. Figure 1 shows a 100-node fixed sensor network in a playing field of size 50m x 50m with the base station (BS) fixed and far away from all the sensor nodes.
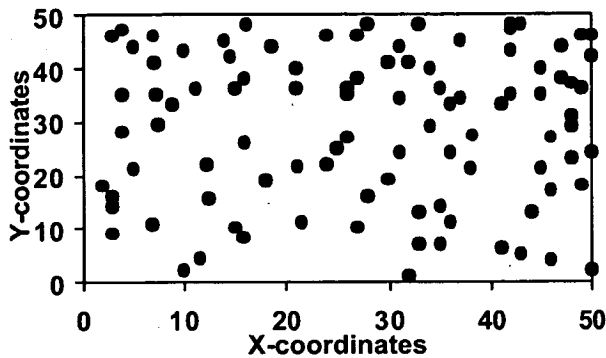
Figure 1. Random 100-node topology for a 50m x 50m network. BS is located at (25, 150), which is at least 100m from the nearest node.

In this paper we assume the following:

- Each sensor node has power control and the ability to transmit to any other node or directly to the BS [5,7].
- Our model sensor network contains homogeneous and energy constrained sensor nodes with uniform energy.
- Every node has location information.
- There is no mobility.

An important operation in a sensor network is systematic gathering of data from the field, where each node has a packet of information in each round of communication [3]. In this operation, data sensed by the nodes need to be combined into a single message and sent to a distant base station. This data fusion among the sensor nodes requires wireless communications. The amount of energy spent in transmitting a packet has a fixed cost in electronics and a variable cost that depends on the distance of transmission. Receiving a data packet also has a similar fixed energy cost in electronics. Therefore, to conserve energy short distance transmissions are preferred. In order to balance the energy spent in the sensor nodes, nodes should take turns transmitting to the BS, as this is an expensive transmission.

In each round of this data-gathering application, all data from all nodes need to be collected and transmitted to the BS, where the end-user can access the data. A simple approach to accomplish this task is for each node to transmit its data directly to the BS. Since the BS is located far away, the cost to transmit to the BS from any node is high, and therefore, the total energy cost per round will be high. In sensor networks, data fusion helps to reduce the amount of data transmitted between sensor nodes and the

BS. Data fusion combines one or more data packets from different sensor measurements to produce a single packet as described in [3]. The LEACH protocol presented in [3] is an elegant solution to this data collection problem, where a small number of clusters are formed in a self-organized manner. A designated node, the cluster head, in each cluster collects and fuses data from nodes in its cluster and transmits the result to the BS. LEACH uses randomization to rotate the cluster heads and improves the energy cost per round by a factor of 4 compared to the direct approach for the 100 node network of Figure 1.

Recently, we developed an improved protocol called PEGASIS (Power-Efficient GAthering in Sensor Information Systems), which requires less energy per round compared to LEACH [6]. The key idea in PEGASIS is to form a chain among the sensor nodes so that each node will receive from and transmit to a close neighbor. Gathered data moves from node to node, get fused, and eventually a designated node transmits to the BS. Nodes take turns to transmit to the BS so that the average energy spent by each node per round is reduced. Building a chain to minimize the total length is similar to the traveling salesman problem, which is known to be intractable. However, with the radio communication energy parameters, a simple chain built with a greedy approach performs quite well. PEGASIS protocol achieves up to 100% improvement with respect to energy cost per round compared to the LEACH protocol. In this paper we will not describe PEGASIS, but evaluate PEGASIS and two new protocols in terms of energy*delay for the data gathering application.

Our schemes can be modified appropriately if some of the stated assumptions about sensor nodes are not valid. If nodes are not within transmission range of each other, then alternative, possibly multi-hop transmission paths will have to be used. In fact, our chain based schemes will not be affected that much as each node communicates only with a local neighbor and we can use a multi-hop path to transmit to the BS. We need to make some adjustments in the chain construction procedure to ensure that no node is left out. Other schemes, including LEACH, rely on direct reach ability to function correctly. To ensure balanced energy dissipation in the network, an additional parameter could be considered to compensate for nodes that must do more work every round. If the sensor nodes have different initial energy levels, then we could consider the remaining energy level for each node in addition to the energy cost of the transmissions. The assumption of location information is not critical. The BS can determine the locations and transmit to all nodes, or the node can determine this through received signal strengths. For example, nodes could transmit progressively reduced signal strengths to find a close neighbor to exchange data. This would require the nodes to consume some energy when trying to find local

1

neighbors, however, this is only a fixed initial energy cost when constructing the chain. If nodes are mobile, then different methods of transmission could be examined. For instance, if nodes could approximate how often and at what speed other nodes are moving, then it could determine more intelligently how much power is needed to reach the other nodes. Perhaps, the BS can help coordinate the activities of nodes in data transmissions. Discussion of schemes with mobile sensor nodes is beyond the scope of this paper.

Another important factor to consider in the data gathering application is the average delay per round. Here, we assume that data gathering rounds are far apart, and the only traffic in the network is due to sensor data. Therefore, data transmissions in each round can be completely scheduled. The delay for a packet transmission (we assume that all packets are 2000 bits long) is dominated by the transmission time as there is no queuing delay and the processing and propagation delays are negligible compared to the transmission time. With the direct transmission scheme, nodes will have to transmit to the base station one at a time, making the delay a total of 100 units (1 unit per transmission). The linear chain-based scheme, although energy efficient, will also require 100 units of delay as the transmissions are sequential. To reduce delay, one needs to perform simultaneous transmissions. The well known approach of using a binary scheme to combine data from N nodes in parallel will take about log N units of delay, although incurring an increased energy cost. *Energy*Delay* is an interesting metric to optimize per round of data gathering.

Simultaneous wireless communications among pairs of nodes is possible only if there is minimal interference among different transmissions. CDMA technology can be used to achieve multiple simultaneous wireless transmissions with low interference. If the sensor nodes are CDMA capable, then it is possible to use the binary scheme and perform parallel communications to reduce the overall delay. However, the energy cost may have to go up slightly as there will still be a small amount of interference from other unintended transmissions. Alternatively, with a single radio channel and non-CDMA nodes, simultaneous transmissions are possible only among spatially separated nodes. Since the energy costs and delay per transmission for these two types of nodes are quite different, we will consider *energy*delay* reduction for our data gathering problem separately for these two cases.

In this paper we present two protocols for *energy*delay* reduction: a chain-based binary combining protocol that uses CDMA capable nodes and a 3 level hierarchy chain-based protocol for non-CDMA nodes. A chain is formed among the sensor nodes in both of these protocols so that each node will receive from and transmit to a close neighbor at the lowest level of the hierarchy. Gathered data move from node to node, get fused, and eventually a designated node transmits to the BS. Nodes take turns transmitting to the BS so that the average energy spent by each node per round is reduced. The binary scheme has a hierarchy of log N, with N equal to the number of nodes. The binary scheme would therefore have a delay of 7 +1 (for transmitting to the base station) for 100 nodes and performs better than LEACH by a factor of 8. The 3 level hierarchy chain-based protocol has a higher delay but is better than the binary scheme with non-CDMA nodes. This is because in the binary scheme there are many nearby simultaneous transmissions at the lower levels and the interference will be very high. In the 3 level scheme, fewer and distant simultaneous transmissions take place causing less interference. This 3 level chain-based protocol performs better than the direct scheme by a factor of about 60.

The paper consists of the following sections. In Section 2, the radio model for energy calculations is discussed. In Section 3, an analysis of the *energy*delay* metric for data gathering is given. Section 4 describes the chain-based binary approach using CDMA capable sensor nodes. Section 5 presents the chain-based 3 level scheme without CDMA capable sensor nodes. Experimental results are given in Section 6. Finally, Section 7 concludes the paper and proposes future work.

2. Radio Model for Energy Calculations

We use the same radio model as discussed in [3] which is the first order radio model. In this model, a radio dissipates E_{elec} = 50 nJ/bit to run the transmitter or receiver circuitry and ϵ_{amp} = 100 pJ/bit/m² for the transmitter amplifier. The radios have power control and can expend the minimum required energy to reach the intended recipients. The radios can be turned off to avoid receiving unintended transmissions.

An r^2 energy loss is used due to channel transmission [8,11]. The equations used to calculate transmission costs and receiving costs for a *k*-bit message and a distance *d* are shown below:

Transmitting
$$E_{Tx}(k, d) = E_{Tx-elec}(k) + E_{Tx-amp}(k,d)$$
$$E_{Tx}(k, d) = E_{elec}*k + \epsilon_{amp} * k* d^2$$

Receiving
$$E_{Rx}(k) = E_{Rx-elec}(k)$$
$$E_{Rx}(k) = E_{elec}*k$$

Receiving is also a high cost operation, therefore, the number of receives and transmissions should be minimal.

In our simulations, we used a packet length *k* of 2000 bits. With these radio parameters, when d² is 500m², the energy spent in the amplifier part equals the energy spent

2

in the electronics part, and therefore, the cost to transmit a packet will be twice the cost to receive.

It is assumed that the radio channel is symmetric so that the energy required to transmit a message from node i to node j is the same as energy required to transmit a message from node j to node i for a given signal to noise ratio (SNR), typically 10 dB. When there are multiple simultaneous transmissions, the transmitted energy should be increased to ensure that the same SNR as with a single transmission is maintained. With CDMA nodes using 64 or 128 chips per bit (which is typical) the interference from other transmissions are calculated as a small fraction of the energy from other unintended transmission. This effectively increases the energy cost to maintain the same SNR. With non-CDMA nodes, the interference will equal the amount of energy seen at the receiver from all other unintended transmitters. Therefore, only few spatially distant pairs can communicate simultaneously.

3. *Energy*Delay Analysis for Data Gathering*

In this section we will analyze the *energy*delay* cost per round for data gathering from a sensor web to the distant BS. Recall that the data collection problem of interest is to send a k-bit packet from each sensor node in each round. Of course, the goal is to keep the sensor web operating as long as possible but minimize delay at the same time. A fixed amount of energy is spent in receiving and transmitting a packet in the electronics, and an additional amount proportional to d^2 is spent while transmitting a packet. There is also a cost of 5 nJ/bit/message for data fusion. The delay cost can be calculated as units of time. On a 2Mbps link, a 2000 bit message can be transmitted in 1ms. Therefore each unit of delay will correspond to about 1 ms time for the case of a single channel and non-CDMA sensor nodes. The actual delay value will be different with CDMA nodes depending on the effective data rate. For each of the systems, we assume that the delay is 1 unit for each 2000 bit message transmitted.

The *energy*delay* cost for data gathering in a network of N nodes will be different for the schemes considered in this paper and will depend on the node distribution in the playing field. Consider the example network where the N nodes are along a straight line with equal distance of d between each pair of nodes and the BS at a faraway distance from all nodes. The direct approach will require high energy cost and the delay will be N as nodes transmit to the BS sequentially. The PEGASIS scheme [6], which is near optimal in terms of energy cost for this data gathering application in sensor networks, forms a chain among the sensor nodes so that each node will receive from and transmit to a close neighbor. For this linear network with equally spaced nodes, the energy cost in

PEGASIS is minimized and the variable cost is proportional to $N*d^2$ and the delay will be N units. Therefore, the *energy*delay* cost will be N^2*d^2.

In the binary scheme with perfect parallel transmission of data, there will be N/2 nodes transmitting data to their neighbors at distance d in the lowest level. The nodes that receive data will fuse the data with their own data and will be active in the next level of the tree. Next, N/4 nodes will transmit data to their neighbors at a distance 2d and this procedure continues until a single node finally transmits the combined message to the BS. Thus, for the binary scheme the energy cost will be

$$N/2 * d^2 + N/4 * (2d)^2 + N/8 * (4d)^2 + ...+1 * (N/2*d)^2$$

since the distance doubles as we go up the hierarchy. In addition, there will be a single transmission to the BS and the energy cost depends on the distance to the BS. Without including this additional cost, simplifying the above expression we get for the energy cost for the binary scheme as

$$N/2 * d^2 * (1 + 2 + 4 + ...+ N/2),$$

which equals

$$N(N-1)/2 * d^2.$$

With the delay cost of about log N units, the *energy*delay* for the binary scheme is $N^2/2*d^2*logN$. Therefore, for this linear network, the binary scheme will be more expensive than PEGASIS in terms of *energy*delay*. For random distribution of nodes in a rectangular playing field, the distances do not double as we go up the hierarchy in the binary scheme, and the reduced delay will help reduce the *energy*delay* cost. It is difficult to analyze this cost for randomly distributed nodes and we will use simulations to evaluate this cost.

For the rest of the analysis, we assume a 100-node sensor network in a square field with the BS located far away. In this scenario, energy costs can be reduced if the data is gathered locally among the sensor nodes and only a few nodes transmit the fused data to the BS. This is the approach taken in LEACH [3], where clusters are formed dynamically in each round and cluster-heads (leaders for each cluster) gather data locally and then transmit to the BS. Cluster-heads are chosen randomly, but all nodes have a chance to become a cluster-head in LEACH, to balance the energy spent per round by each sensor node. Nodes are able to transmit simultaneously to their cluster-heads using CDMA. For a 100-node network in a 50m x 50m field with the BS located at (25,150), which is at least 100m from the closest node, LEACH reduces the *energy*delay* cost compared to the direct scheme. For the linear network of N nodes that are equally spaced,

3

LEACH will have slightly higher energy compared to PEGASIS due to the cluster heads transmissions to the BS and a delay of roughly N/c where c is the number of clusters. With 5 clusters suggested in [3], the *energy*delay* for LEACH will be lower than for PEGASIS for a 50m x 50m network. However, for a 100m x 100m network, the *energy*delay* for LEACH will be higher than for PEGASIS.

4. A Chain-based Binary Approach using CDMA

First, we consider a sensor network with nodes capable of CDMA communication. With this CDMA system, it is possible for node pairs that communicate to use distinct codes to minimize radio interference. Thus, parallel communication is possible with 50 pairs for the 100-node network of interest. In order to minimize the delay, we will combine data using as many pairs as possible in each level which results in a hierarchy of log N levels. At the lowest level, we will construct a linear chain among all the nodes, as was done in PEGASIS, so that adjacent nodes on the chain are nearby. For constructing the chain, we assume that all nodes have global knowledge of the network and employ the greedy algorithm. The greedy approach to constructing the chain works well, and this is done before the first round of communication. To construct the chain, we start with the furthest node from the BS. We begin with this node in order to make sure that nodes farther from the BS have close neighbors. As in the greedy algorithm the neighbor distances will increase gradually since nodes already on the chain cannot be revisited.

For gathering data in each round, each node transmits to a close neighbor in a given level of the hierarchy. This occurs at every level in the hierarchy, but the only difference is that the nodes that are receiving at each level are the only nodes that rise to the next level. Finally, at the top level the only node remaining will be the leader, and the leader will transmit the 2000 bit message to the BS. Note that node i will be in some random position j on the chain. Nodes take turns transmitting to the BS, and we will use node number i mod N (N represents the number of nodes) to transmit to the BS in round i. In Figure 2, for round 3, node c3 is the leader. Since, node c3 is in position 3 (counting from 0) on the chain, all nodes in an even position will send to their right neighbor. Now at the next level, node c3 is still in an odd position so again, all nodes in an even position will fuse its data with its received data and send to their right. At the third level, node c3 is not in an odd position, so node c7 will fuse its data and transmit to c3. Finally, node c3 will combine its current data with that received from c7 and transmit the message to BS. The chain-based binary scheme performs data fusion at every node that is transmitting except the

end nodes in each level. Each node will fuse its neighbor's data with its own to generate a single packet of the same length and then transmit that to the next node. In the above example, node c0 will pass its data to node c1. Node c1 fuses node c0's data with its own and then transmits to node c3 in the next level. In our simulations, we ensure that each node performs equal number of sends and receives after N rounds of communication, and each node transmitting to the BS in one of N rounds. We then calculate the average energy cost per round, while the delay cost is the same for each round.

BS

$\uparrow$

c3

c3 $\leftarrow$ c7

c1 $\rightarrow$ c3 c5 $\rightarrow$ c7

c0 $\rightarrow$ c1 c2 $\rightarrow$ c3 c4 $\rightarrow$ c5 c6 $\rightarrow$ c7
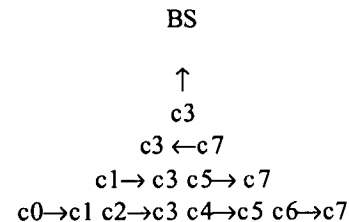
Figure 2. Data gathering in a chain-based binary scheme.

The chain-based binary scheme improves on LEACH by saving energy and delay in several stages. At the lower levels, nodes are transmitting at shorter distances compared to nodes transmitting to a cluster head in the LEACH protocol, and only one node transmits to the BS in each round of communication. By allowing nodes to transmit simultaneously, the delay cost for the binary scheme decreases from that of LEACH by a factor of about 3. While in LEACH, only 5 groups can transmit simultaneously, here at each level, we have multiple nodes transmit simultaneously. At each level of the binary scheme, transmissions are simultaneous making the total delay log N +1 for transmitting to the BS. In LEACH, the delay for 100 node networks will be 27 units. The delay for all nodes to transmit to the cluster-head is the max number of nodes in any of the 5 clusters. If all the clusters are of the same size, then the delay would be 19. Then all 5 cluster-heads must take turns to transmit to the BS, making that a total of 24. For overhead calculations, we have 1 unit of delay for cluster formation, 1 unit of delay for all nodes to broadcast to the cluster-head its presence in that cluster, and finally 1 unit of delay for the cluster-head to broadcast a TDMA schedule to the nodes so that nodes will know when to broadcast to the cluster-head.

4

5. A Chain-based 3 Level Scheme without CDMA

CDMA may not be applicable for all sensor networks as these nodes can be expensive. Therefore, we need a protocol that will achieve a minimal *energy*delay* with non-CDMA nodes. It will not be possible to use the binary scheme in this case as the interference will be too much at lower levels. We either have to increase the energy cost significantly or take more time steps at lower levels of the hierarchy both of which will lead to much higher *energy*delay* cost. Therefore, in order to improve *energy*delay* we need a protocol that allows simultaneous transmissions that are far apart to minimize interference while achieving reasonable delay cost.

Based on our experiments, we suggest the chain-based 3 level scheme for data gathering in sensor networks with non-CDMA nodes. In the 3 level scheme also, we start with the linear chain among all the nodes and divide them into 10 groups. In the 100-node network, therefore, only 10 simultaneous transmissions take place at the same time, and data fusion takes place at each node (except the end nodes in each level). The transmissions are also far enough apart that there is minimal interference. Figure 3 shows an example of this scheme with 100 nodes. Here we would have 10 groups of 10. We will have a different leader each round transmit to the BS to evenly distribute the work load among the sensor nodes. We find the index *i* which will represent the leader position modulo 10. In Figure 3, c18 is our leader. Then all nodes will send their data in the direction of index 8 within their group since 18 modulo 10 is 8. The delay at the first level is 9 units. Then the second level will contain nodes c8, c18, c28...c98. These 10 nodes will be divided into two groups. Since the leader position is 18, all nodes that are in the first group will send down the chain 10 positions from its own position on the chain. So node c48 will send to node c38, and node c38 will send to node c28 and so on. Since node c8's position is less than node c18's, node c8 will transmit to a position that is 10 greater than its own. In group 2, nodes know in which direction to send the data using the leader position + 50. So here, the nodes in group 2 would send in the direction of node c68 in the same manner as in group 1. This gives us a delay of 4 units for the second level. In the third level, node c68 transmits to node c18, and then finally node c18 transmits to the base station, giving us a total delay of 15 units. The transmission schedule can be programmed once at the beginning so that all nodes know where to send data in each round of communication.
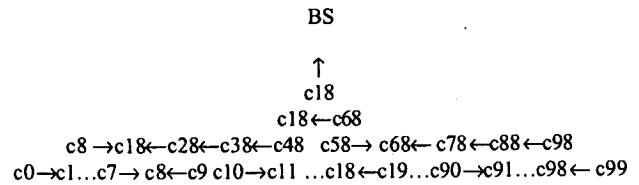
BS

↑

c18

c18←c68

c8 →c18←c28←c38←c48 c58→ c68← c78←c88←c98

c0→c1...c7→ c8←c9 c10→c11 ...c18←c19...c90→c91...c98← c99

Figure 3. Chain-based 3 level scheme for a sensor network with non-CDMA nodes.

6. Experimental Results

To evaluate the performance of the chain-based binary scheme and the chain-based 3 level scheme, we simulated direct transmission, PEGASIS, LEACH, and the two new schemes using several random 100-node networks with CDMA nodes and non-CDMA nodes. The BS is located at (25, 150) in a 50m x 50m field, and the BS is located at (50,300) in a 100m x 100m field. We ran the simulations to determine the energy cost for all the schemes after all 100 nodes had a chance to become leader, with each node having the same initial energy level. We then used these costs to determine the average energy cost per round of data gathering. In both CDMA and non-CDMA systems, we included the interference costs when there are simultaneous transmissions to ensure that the same SNR of 10 dB is maintained as with single transmission. Our simulations show:

- The chain-based binary scheme is approximately 8x better than LEACH and 130x better than direct for a 50m x 50m network in terms of *energy* delay* for sensor networks with CDMA nodes.
- The chain-based binary scheme is approximately 12x better than LEACH and 280x better than direct for a 100m x 100m network in terms of *energy* delay* for sensor networks with CDMA nodes.
- The chain-based 3 level scheme is approximately 4x better than PEGASIS and 60x better than direct for a 50m x 50m network in terms of *energy* delay* for sensor networks with non-CDMA nodes.
- The chain-based 3 level scheme is approximately 4x better than PEGASIS and 140x better than direct for a 100m x 100m network in terms of *energy* delay* for sensor networks with non-CDMA nodes.
- A more balanced energy dissipation among the sensor nodes to have full use of the complete sensor network.

These results are summarized in Table 1 and Table 2.

5

Table 1. *Energy* delay* cost for Direct, PEGASIS, LEACH, chain-based binary scheme and the chain-based 3 level scheme. These results are for a 50m x 50m network.

Protocol	Energy	Delay	Energy* Delay
Direct (both systems)	0.32993	100	32.9938
PEGASIS (both systems)	0.024008	100	2.4008
LEACH (CDMA nodes)	0.079696	27	2.1518
Chain-based binary (CDMA nodes)	0.031847	8	0.2547
Chain-based 3 level (non-CDMA nodes)	0.035772	15	0.5365

Table 2. *Energy* delay* cost for Direct, PEGASIS, LEACH, chain-based binary scheme and the chain-based 3 level scheme. These results are for a 100m x 100m network.

Protocol	Energy	Delay	Energy *Delay
Direct (both systems)	1.280459	100	128.0459
PEGASIS (both systems)	0.036107	100	3.6107
LEACH (CDMA nodes)	0.204786	27	5.5292
Binary (CDMA nodes)	0.055898	8	0.4516
3 Level (non-CDMA nodes)	0.058287	15	0.8743

7. Conclusions and Future Work

In this paper, we describe two new protocols for *energy*delay* reduction for data gathering in sensor networks -- a chain-based binary scheme for sensor networks with CDMA nodes and a chain-based 3 level scheme for sensor networks with non-CDMA nodes. The binary scheme performs better than direct, PEGASIS, and LEACH. It performs better than LEACH by a factor of about 8, about 10 times better than PEGASIS, and more than 100 times better when compared to the direct scheme. In these experiments, the interfering transmissions contribute 1/128 the value of their transmission energy. The chain-based 3 level scheme with non-CDMA nodes outperforms PEGASIS by a factor of 4 and is better than direct by a factor of 60. This chain based scheme outperforms PEGASIS by dividing the chain into "groups" and allowing simultaneous transmissions among pairs in different groups. While energy is still minimal, the delay is decreased from 100 units to 15 units.

It is not clear as to what is the optimal scheme for minimizing *energy*delay* in a sensor network. Since the energy costs of transmissions depend on the spatial distribution of nodes, there may not be a single scheme that is optimal for all sizes of the network. Our preliminary experimental results indicate that for all small networks, the binary scheme performs best as minimizing delay achieves best result for *energy*delay*. With larger networks, we expect that nodes in the higher levels of the hierarchy to be far apart and it is possible that a different multi-level scheme may outperform the binary scheme. When using non-CDMA nodes, interference effects can be reduced by carefully scheduling simultaneous transmissions. Since there is an exponential number of possible schedules, it is intractable to determine the optimal scheduling to minimize *energy*delay* cost. A practical scheme to employ will depend on the size of the playing field and the distribution of nodes in the field.

In this paper, we restricted our discussions to the d^2 model for energy dissipation for wireless communications. In our future work, we will consider higher order energy dissipation models and develop schemes to minimize *energy*delay* cost for data gathering application.

References

[1] "Bluetooth Initiative," http://www.bluetooth.com, 1999.

[2] D. Estrin, R. Govindan, J. Heidemann, and Satish Kumar, "Next Century Challenges: Scalable Coordination in Sensor Networks," In *Proceedings of Mobicom' 99*, 1999.

[3] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," In *Proceedings of the Hawaii Conference on System Sciences*, Jan. 2000.

[4] J. Kulik, W. Rabiner, and H. Balakrishnan, "Adaptive Protocols for Information Dissemination in Wireless Sensor Networks," In *Proceedings of Mobicom' 99*, 1999.

[5] N. Bambos, "Toward Power-Sensitive Network Architectures in Wireless Communications: Concepts, Issues, and Design Aspects," IEEE Personal Communications, June 1998, pp. 50-59.

[6] S. Lindsey, C. S. Raghavendra, "PEGASIS: Power Efficient GAthering in Sensor Information Systems," Submitted to ICC 2001.

[7] R. Ramanathan and R. Hain, "Topology Control of Multihop Wireless Networks Using Transmit Power Adjustment," In *Proceedings Infocom 2000*, 2000.

[8] T. S. Rappaport, *Wireless Communications*, Prentice-Hall, 1996.

[9] S. Singh, M. Woo and C.S. Raghavendra, "Power-Aware Routing in Mobile Ad Hoc Networks," *In Proceedings ACM/IEEE Mobicom'98*, 1998.

[10] K. M. Sivalingam, J. Chen, P. Agrawal and M. Srivastava, "Design and Analysis of Low-power Access Protocols for Wireless and Mobile ATM Networks," *ACM/Baltzer Wireless Networks* vol. 6, pp. 73-87, Feb. 2000.

[11] R. Steele, *Mobile Radio Communications*, Pentech Press, London, 1992.

[12] M. Stemm, P. Gauthier, D. Harada and R. Katz, "Reducing Power Consumption of Network Interfaces in Hand-Held Devices," In *Proceedings 3rd Intl. Workshop on Mobile Multimedia Communications*, Sept. 25-27, 1996, Princeton, NJ.

[13] The WINS Project, http://www.janet.ucla.edu/WINS

[14] M. Zorzi and R. R. Rao, "Energy Management in Wireless Communications," In *Proceedings 6th WINLAB Workshop on Third Generation Wireless Information Networks*, March 1997.

7

Google `bs transmit index direction position location si¢` **Search Patents**

Communication system with a beamformed control channel and method of system control Paul Crichton et al

Patent summary
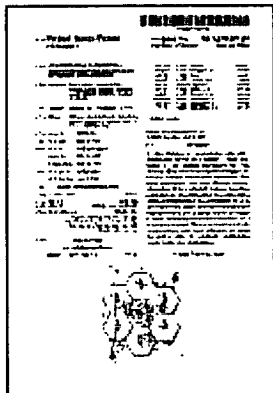
| Read this patent |

View patent at USPTO

Abstract
To reduce interference in a communication system (10), a communication unit (42-50) is arranged to initiate establishment of a radio frequency communication with a base station (26-38) by transmitting a system access request on a dedicated, wide area control channel. Upon receipt of the system access request, a base station (32) of the communication system of FIG. 1 responds by forming a narrowbeam control channel to the communication unit and transmitting system control information to the communication unit on the narrowbeam control channel, the system control information transmitted from the array of antenna elements and arranged to identify a narrowbeam communication resource for use in the radio communication. The communication unit (42-50), upon receiving the system control information, then configures itself to utilise the narrowbeam communication resource for the radio communication.

Patent number: 6330459
Filing date: Mar 12, 1999
Issue date: Dec 11, 2001
Inventors: Paul Crichton, Paul Howard, Nicholas Anderson, William Robinson
Assignee: Motorola, Inc.
Primary Examiner: Jean A Gelin

Claims

What is claimed is:

1. A method of establishing radio communication between a communication unit and a base station having an array of antenna elements, the method comprising the steps of:

receiving a system access request at each base station of a plurality of base stations, wherein each base station of the plurality of base stations comprises an antenna array; making signal parameter measurements of the access request received by each base station of the plurality of base stations; determining a rank order of signal parameter measurements with respect to the plurality of base stations; selecting a base station to serve the communication unit based on the rank order to produce a serving base station; at the serving base station, in response to the received system access request, forming a first narrowbeam control channel to the communication unit and transmitting system control information to the communication unit on the first narrowbeam control channel, the system control information transmitted from the array of antenna elements and arranged to identify a narrowbeam communication resource for use in the radio communication; at the serving base station, receiving a request for assignment of a new narrowbeam control channel; in response to the request for assignment of a new narrowbeam control channel, instructing at least one non-serving base station of the plurality of base stations to prepare to transmit narrowbeam control channels at the communication unit; at the at least one non-serving base station of the plurality of base stations, in response to receiving an instruction to prepare to transmit, notifying the serving base station of channel assignment information pertaining to a subsequent transmission of the narrowbeam control channels at the communication unit; and at the serving base station, in response to receiving the channel assignment information from the at least one non-serving base station, notifying the communication unit of the channel assignment information on the first narrowbeam control channel.

Current U.S. Classification
455/562; 455/434; 455/436

International Classification
H04Q 720

Search within this patent

[   ] [Search]

Citations

Patent Number	Title	Issue date
5203010	Radio telephone system incorporating multiple time periods for communication transfer	Apr 13, 1993
5487174	Methods in a cellular mobile radio communication system	Jan 23, 1996
5628052	Wireless communication system using distributed switched antennas	May 6, 1997
5649293	Method for assigning subscribers between narrowbeam sectors	Jul 15, 1997
5710982	Power control for TDMA mobile satellite communication system	Jan 20, 1998
5887261	Method and apparatus for a radio remote repeater in a digital cellular radio communication system	Mar 23, 1999
5907809	Position determination	May 25, 1999

2. The method of claim 1, further comprising the step of, at the serving base station, periodically altering a beam pattern of the first narrowbeam control channel.

3. The method of claim 2, wherein the beam pattern of the first narrowbeam control channel is oscillated about an expected position of the communication unit.

4. The method of claim 2, wherein a width of the beam pattern of the first narrowbeam control channel is pulsed.

5. The method of claim 2, further comprising the step of, at the serving base station, transmitting beam pattern information on the first narrowbeam control channel identifying how the beam pattern of the first narrowbeam control channel is altering.

6. The method of claim 1, further comprising the steps of:

determining a base station handoff candidate from among the at least one non-serving base stations; and
receiving, by the base station handoff candidate and via a narrowbeam control channel associated with the base station handoff candidate, a signal to initiate handoff of the radio communication.

7. The method of claim 1, further comprising the steps of:

storing base station location information and communication unit location information; and
by each base station of the plurality of base stations, accessing the base station location information and the communication unit location information to beamform, during handoff of the radio communication, a narrowbeam control channel in a direction of the communication unit.

8. A radio communication system for supporting radio communication between a communication unit and at least one of a plurality of base stations, the system comprising:

a plurality of base stations, wherein each base station of the plurality of base stations comprises:
    an array of antenna elements;
    a means, responsive to the array of antenna elements, for receiving and processing a system access request;

	using multiple base station signals	
5907816	High gain antenna systems for cellular use	May 25, 1999
5960350	Method and system for optimizing a traffic channel in a wireless communications system	Sep 28, 1999
6091788	Base station equipment and a method for steering an antenna beam	Jul 18, 2000
6131034	Method and apparatus for collector arrays in wireless communications systems	Oct 10, 2000

Referenced by

Patent Number	Title	Issue date
6445917	Mobile station measurements with event-based reporting	Sep 3, 2002
6701164	Mobile communication system with zone-shift control apparatus	Mar 2, 2004
6771987	Method and apparatus for uplink scheduling	Aug 3, 2004
6889047	Wireless base station and a wireless phone	May 3, 2005
6947749	Apparatus for increasing cell capacity in mobile communication system using adaptive sectorization	Sep 20, 2005

a means, responsive to the system access request, for assigning and generating system control information identifying a narrowbeam communication resource for use in a radio communication with the communication unit;
a means, coupled to the array of antenna elements, for forming and transmitting a first narrowbeam control channel to the communication unit in response to the received system access request;
a means for making signal parameter measurements of the access request received at each base station of the plurality of base stations;
a means for determining a rank order of signal parameter measurements with respect to the plurality of base stations;
a means for selecting a serving base station to serve the communication unit from the rank order;
a means at the serving base station for receiving a request for assignment of a new narrowbeam control channel;
a means for instructing, in response to the request for assignment of a new narrowbeam control channel, at least one non-serving base station of the plurality of base stations to prepare to transmit narrowbeam control channels at the communication unit;
a means at the at least one non-serving base station of the plurality of base stations for notifying, in response to receiving an instruction to prepare to transmit, the serving base station of channel assignment information pertaining to a subsequent transmission of the narrowbeam control channels at the communication unit; and
a means at the serving base station for notifying, in response to receiving the channel assignment information from the at least one non-serving base station, the communication unit of the channel assignment information on the first narrowbeam control channel.

9. The communication system of claim 8, wherein the serving base station further comprises means for periodically altering a beam pattern of the first narrowbeam control channel.

Drawings

more »

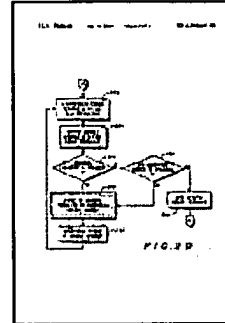|		
	and method for controlling the same	
6970708	System and method for improving channel monitoring in a cellular system	Nov 29, 2005
6996418	Apparatus and method for OFDM data communications	Feb 7, 2006
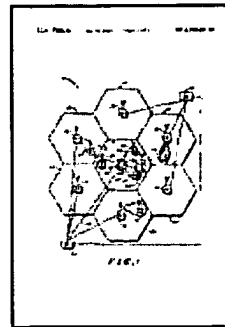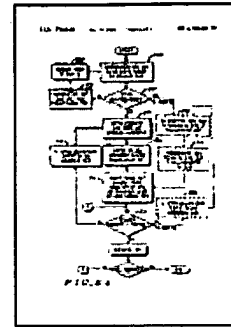
Page 2

Page 3

Page 4

Google Home - About Google - About Google Patent Search

©2007 Google

Resource Reservation and Allocation Based on Direction Prediction for Handoff in Mobile Multimedia Networks

Jongchan Lee[1], Hongjin Kim[2], and Kuinam J. Kim[3]

[1] Radio Access Research Team ETRI, Korea
chan2000@etri.re.kr
[2] Dept. of Computer Information KyungWon College, Korea
[3] Dept. of Information Security Konggi Univ., Korea

Abstract. Future mobile communication systems can support not only voice but also multimedia applications such as data, image and video. It requires greater resources than voice-oriented mobile system. If handoff events are occurred during the transmission of multimedia, the efficient resource allocation and handoff procedures are necessary to maintain the same QoS of transmitted multimedia traffic because the QoS may be defected by some delay and information loss. This paper proposes a resource reservation and allocation scheme to accommodate multimedia traffic based on the direction estimation in mobile multimedia networks. This scheme estimates the position of mobiles based on a two step estimation comprised of sector estimation, zone estimation. With the position information, the moving direction is determined.

1 Introduction

The explosive growth of Internet access in parallel with the technological advances in mobile communications has motivated mobile computing and multimedia applications in wireless mobile networks. A Key characteristic of multimedia services is that they require different Quality of Service (QoS) guarantees. Due to the limitations of the radio spectrum, the wireless systems use micro-cellular architectures in order to provide a higher capacity. Because of small coverage area of micro-cells, network resources availability varies frequently as users move from one access point to another [1]. In order to deterministically guarantee QoS support for a mobile unit, the network must have prior exact knowledge of the mobile's mobility. Majority of the existing schemes to support mobility make a reservation for resources in adjacent cells. However these techniques cause a waste of resources since it is regardless of the direction of mobiles. Also, existing methods for predicting and reserving resources for future handoff calls do not seem to be suitable for mobile multimedia networks. The amount of resources required to successfully perform handoff may vary arbitrarily over a wide range in a mobile multimedia network. For example, data and video applications may adapt to different service quality levels and consequently

may accept different levels of resources in order to ensure a successful handoff [2][3][4]. In this paper, we consider a mobile network supporting diverse traffic characteristics of voice, data, and video applications. Since the connections can now differ in the amount of resources required to meet their QoS needs, the question is how should a base station dynamically adapt the amount of resources reserved for dealing with handoff requests. In this paper, the handoff requests for real-time connections are handled based on the direction prediction and the resource reservation scheme. The resources in the estimated adjacent cells should be reserved to guarantee the continuity of the real-time connections. If handoff requests are occurred during the transmission of multimedia traffic, the efficient resource allocation and handoff procedures are necessary to maintain the same QoS of transmitted multimedia traffic because the QoS may be defected by some delay and information loss. This paper proposes a handoff scheme to transmit multimedia traffic based on the resource reservation using direction estimation.

2   Direction Prediction Method

Figure 1 shows how our scheme divides a cell into many zones based on the signal strength, and then estimates the optimal zone stepwise where the mobile is located. This process is based on a two step location estimations which determines the mobile position by gradually reducing the area of the mobile position [5][6]. This scheme is implemented as an estimator into the base station. The estimator is started with a timer, and then the estimation is performed sequentially in two steps. The estimator first estimates the location sector in the sector estimation step, then estimates the location zone in the zone estimation step, and then finally estimates mobile's direction.

2.1   Sector Estimation

The sector estimation, the first step of the location estimation, is done in the following procedure.

A. All the neighboring base stations transmit pilot signals periodically.
B. The demodulator of the mobile measures PSSs of neighboring base stations.
C. The mobile sends PSMM (Pilot Strength Measurement Message) to the base station.
D. The estimator in the base station compares the received strengths of the pilot channels with each other and chooses the sector neighboring to the base station of the greatest signal strength as the sector at which the mobile locates.
E. The sector number is registered to the object information.
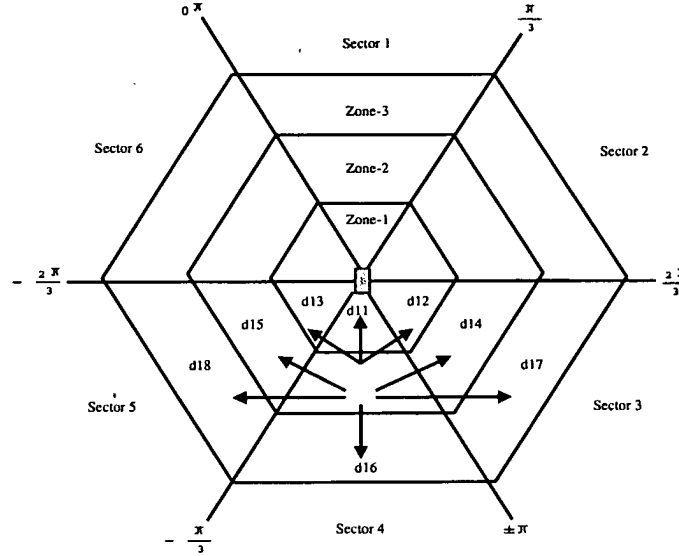
Fig. 1. Sector, zone and a mobile's moving direction

2.2   Zone Estimation

Each cell is divided into n zones with each zone classified by PSS. The following LOS algorithm summarizes the zone estimation procedure for LOS model.

1. Select each threshold considering PSS.
2. In order to map the signal strength onto the direction information, determine the distance function for each threshold with Equation(1).

$$P_A(d) = k_1 - k_2(d) + u(t)$$
$$P_B(d) = k_1 - k_2(D - d) + v(t) \qquad (1)$$

In Equation (1) D indicates the distance between two base stations, and d the distance between the base station A and the mobile. k_1 is proportional to the transmission power of the station and k has the offset value depending the radio propagation environment. Two random signals $u(t)$, $v(t)$ which indicate the power distributions of signals received at the distance d respectively from the station A and from the station B have i.i.d (identical independent distribution) with Gaussian distribution of $N(\mu(d), \sigma)$. The average value of the received signal at the specific location, $\mu(d)$, is determined by the path-loss component proportional to the distance and σ is assumed to be same. The changes in the LOS are depicted by k_2.

3. Classify zones using the distance function.
4. Assign the zone number and the PSS threshold to all divided zones.

Using the sector of blocks selected in the sector estimation, the zone estimation, the second step of the location estimation, estimates the zone of blocks at one of which the mobile locates. It is done in the following procedure.

A. The base station transmits the pilot signal periodically.
B. The demodulator of the mobile measures the signal strength of the pilot channel of the base station in which it is.
C. The mobile sends PSMM to the base station.
D. The estimator estimates the zone using the LOS algorithm.
E. The estimated zone number is registered to a zone object.

2.3   Direction Estimation

The estimator estimates a sector in the sector estimation step and a zone in the zone estimation step respectively, and then finally computes mobile's direction using the vector information between the estimated location and the previous one. In order to indicate the location of each zone within a cell, we use the vector data which is obtained by converting the rectangular coordinate of the zone to the polar coordinate with the origin of the base station [5]. Each vector has the information on a distance and an angle. The polar coordinate indicates the location by the distance from the origin to the mobile and the angle from the positive horizontal axis toward counter-clockwise. In our study we need the direction information to identify the sector relative to the base station and the relative position of a zone so we use the polar coordinates converted from the rectangular coordinates.

A mobile's direction is classified based on the movement into the upper zone (d11, d12 and d13), lower one (d16, d17 and d18) and the same one (d14 and d15), as shown in Figure 1. The moving radius of a mobile toward the upper from the lower zone becomes wider, while the moving radius of a mobile toward the lower zone becomes narrower. That is, the direction prediction for a mobile toward his BS is difficult more than mobiles toward his cell boundary. Therefore, for a mobile toward the upper zone it is efficient to increase the number of cells within his moving radius that is expected to be handed off, and to decrease the number of cells in case of a mobile toward the lower zone. Figure 2 shows the direction estimation algorithm based on the above conditions.

2.3.1   Resource Reservation for Low-Speed Mobiles

The moving radius and the moving pattern of a mobile has different characteristics according to the speed of the mobile. That is, a low-speed mobile (a pedestrian) has a smaller moving radius and a more complex moving pattern, while a high-speed mobile (a motor vehicle) has a larger radius and a simpler pattern. Using those characteristics, reservation variables such as the current position and the moving direction of a mobile are defined, and the neighboring cells that need to reserve resources are decided. In the case of low-speed mobiles,
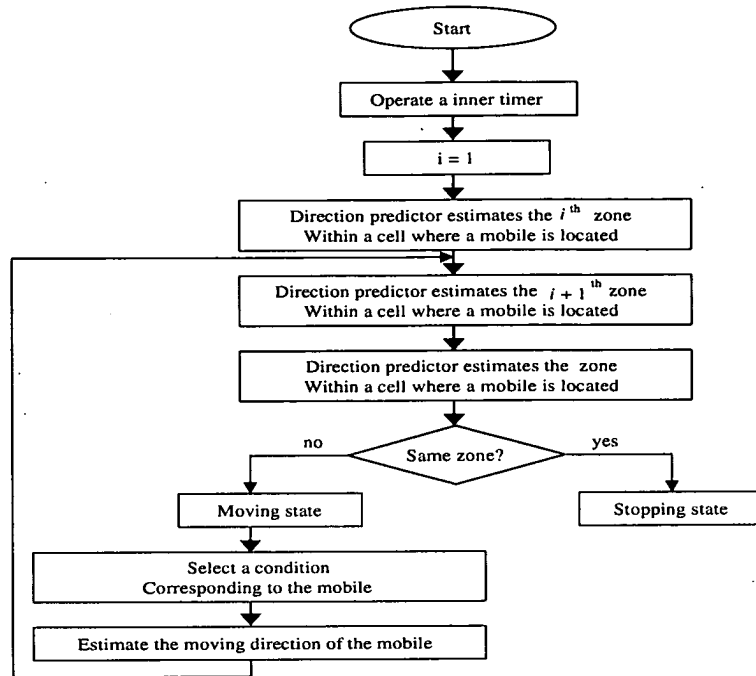
Fig. 2. Direction estimation algorithm

the result of the direction estimation can vary significantly due to two factors. First, the current position is considered. This is because the moving radius of low-speed mobiles is narrow, therefore, whether the handoff is done or not can be estimated according to his current position. Second, the moving direction is considered. The reason of using this factor is that the handoff attempt of the mobiles moving toward the inside of cell is decreased, while the handoff attempt of the mobiles moving toward the outside of cell is increased. The location zone is estimated in the sector and the zone estimation, and then based on the location of the estimated zone and the predicted direction, cells in the neighborhood are ranked according to the likelihood that the mobile will move into these cells. Cells needed to reserve the resource is decided by the sector estimation, and the resource reservation for the cells is done using the current position estimated by the zone estimation. Mobiles moving toward the upper zone need not reserve the resource. The resources for mobiles moving toward the lower zone are reserved only in case their estimated position is zone-3. Resource reservation conditions based on the reservation variable for a low-speed mobile is as follows.

– Condition 1: if the current position is zone-1, the reservation is not made regardless of its moving direction.

- Condition 2: if the current position is zone-2 and its movement is done from zone-1, the reservation is not made.

- Condition 3: if the current position is zone-2 and its movement is done from zone-3, the reservation is made to the maximum two cells.

- Condition 4: if the current position is zone-3 and its movement is done from zone-2, the reservation is made to the maximum one cell.

2.3.2 Resource Reservation for High-Speed Mobiles

The reservation variable for fast-speed mobiles is the moving direction. If the mobiles move fast, the feasibility of performing a handoff is expected to be higher; therefore the reservation is needed regardless of their current position within cell. Since the mobility of the fast-speed mobiles has a varying randomness, a cluster of cells that is reflective of the user mobility is needed to reserve the resources. Resource reservation conditions based on the reservation variable for a high-speed mobile is as follows.

- Condition 1: If the mobile moves from zone-1 to zone-2 within the same sector, the resources are reserved for three cells that is reflective of the moving direction.
- Condition 2: If the mobile moves from zone-2 to zone-3 within the same sector, the resources are reserved for one cell that is reflective of the moving direction.
- Condition 3: If the mobile moves from zone-3 to zone-2 within the same sector, the resources are reserved for five cells that is reflective of the moving direction.
- Condition 4: If the mobile moves from zone-1 to zone-2 within the same sector, the resources are reserved for three cells that is reflective of the moving direction.
- Condition 5: If the mobile moves from zone-1 to zone-2 within the other sector, the resources are reserved for three cells that is reflective of the moving direction.
- Condition 6: If the mobile moves from zone-2 to zone-3 within the other sector, the resources are reserved for one cell that is reflective of the moving direction.
- Condition 7: If the mobile moves from zone-3 to zone-2 within the other sector, the resources are reserved for five cells that is reflective of the moving direction.
- Condition 8: If the mobile moves from zone-2 to zone-1 within the other sector, the resources are reserved for five cells that is reflective of the moving direction

3   Direction Prediction Based Resource Reservation and Allocation

3.1   Resource Reservation and Allocation Structure

A real-time mobile performs resource reservation for hand-off, and a set of the reserved resources can be occupied temporally by non-real-time mobiles within the target cell. On the order hand, a non-real-time mobile doesn't perform resource reservation for handoff, and its resource allocation request is buffered in the waiting queue of the target BS during handoff duration time and is given the priority based on the service demand time. If the reserved set is returned because the corresponding real-time mobile is handed off, the priority of the request becomes the lowest lank. This strategy is explained in Figure 3.
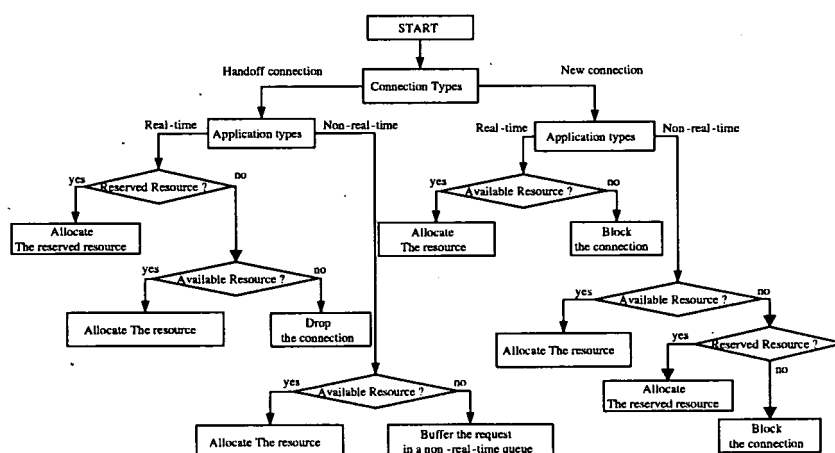
Fig. 3. Admission control for multimedia connections

3.2   Resource Reservation Procedure

The base station reserves only the resources corresponding to the minimum transmission rate to the mobile. Based on the location and the direction of the mobile within a cell, the resource reservation is performed with the following order: unnecessary state, not necessary state, necessary state, and positively necessary state. If the reservation variable for the mobile is changed, the reservation is canceled and the resources have to be released with the reverse order and returned to the fool of available resource. The set of the reserved resources have its priorities depending on whether it can be allocated to new connections or not: a real-time handoff connection (priority 1), a non-real-time handoff connection (priority 2) and a non-real-time new connection (priority 3).

- *Unnecessary State*

  The resource reservation needs not be performed.

  This state corresponds to the resource reservation condition 1 and 2 for low-speed mobiles.

- *Not Necessary State*

  A set of the reserved resources corresponds to priority 3.

  If any resources are available in each of the estimated cells, the resources are then reserved for each of the mobiles. A set of the reserved resources can be occupied by the new connections if enough resources are not available for a new connection in each of the estimated cells.

  If there are any resources available to support the reservation in the estimated cell, and a moving connection competes with a new connection for the resources, the resources are occupied with the following order: a real-time handoff connection, a real-time new connection, a non-real-time handoff connection, and a non-real-time new connection.

  If no resources are available, the reservation is not done.

  This state corresponds to the resource reservation condition 3, 7 and 8 for fast-speed mobiles.

- *Necessary State*

  A set of the reserved resources corresponds to priority 1.

  If there is no enough resource available to accommodate a new connection, a set of the reserved resources for real-time handoff connections can be occupied by non-real-time new connections.

  If there are resources available to support the reservation in the estimated cell, and a moving connection competes with a new connection for the resources, the order of occupying the resources is the same as Not Necessary State.

  If no resources are available for the reservation in the estimate cell, the shared part resources are allocated and reserved for a real-time connection.

  This state corresponds to the resource reservation condition 4 for low-speed mobiles.

  This state corresponds to the resource reservation condition 1, 4, 5 and 8 for fast-speed mobiles.

- *Positively Necessary State*

  The reserved resources correspond to priority 1.

  New connections cannot occupy the reserved resources.

  In case of a moving connection competes with a new connection for resources in the estimated cell, the resources are occupied with the following order: a real-time handoff connection, a non-real-handoff connection, a real-time new connection and a non-real-time new connection.

  If no resources are available for the reservation in the estimate cell, the shared part resources can be allocated and reserved for both real-time connections and non- real-time connections.

This state corresponds to the resource reservation condition 4 for low-speed mobiles.

This state corresponds to the resource reservation condition 2 and 6 for fast-speed mobiles.

4   Simulation Model and Result

The proposed scheme is compared with two different methods to evaluate the performance.

Method 1: there is resource reservation. The resources are reserved exclusively for handoff connections in each cell. The remaining resources can be equally shared among handoff and new connections. This method is called Fixed_Res.

Method 2: the resources are reserved dynamically based on the current connections in the neighboring cells. This method is called Dynamic_Res.
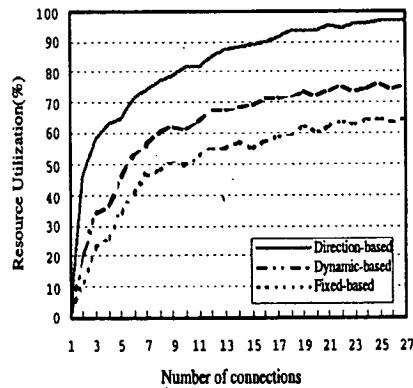
Fig. 4. Comparison of resource utilization
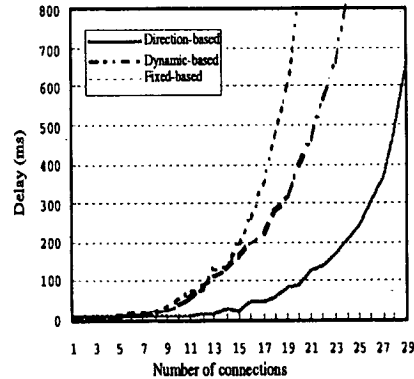
Fig. 5. Comparison of transmission delay

The simulation model composed of a single cell, which will keep contact with its six neighboring cells. Each cell contains a base station, which is responsible for the connection setup and tear-down of new applications and to serve handoff applications. We consider the following simulation parameters regarding the received signal strength. The mean signal attenuation by the path-loss is proportional to 3.5 times the propagation distance, and the shadowing has a log-normal distribution with a standard deviation of $\sigma = 6dB$. A value of the received signal strength less than -16 dB is regarded as an error, which is therefore excluded from the calculation. Figure 4 shows the results for an average percentage of resource utilization as a function of connections arrival with priority to handoff connections over new connections. The resource utilization for
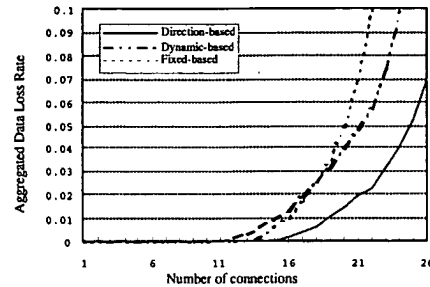
Fig. 6 Comparison of aggregated data loss rate

the Direction-based is increased up to 25 arrival, than that for Fixed and Dynamic. This improvement may be caused since the Direction-based allocates the reserved resources not only real-time handoff connections but also non-real-time new connections. In Figure 5, the comparison of transmission delay of the three schemes is plotted against the number of connections. From the figure, we can see the performance of the direction-based is up to 2.5 times better than that of the conventional schemes. This is because non-real-time connections can adaptively occupy the reserved resources for real-time connections, and we can prevent the performance degradation due to queueing of non-real-time connections.

Figure 6 shows the aggregated data loss rate. It is observed that as the number of connections increase, direction-based provides a noticeable improvement over the conventional schemes for real-time connections, while slightly degrading the performance for the non-real-time connections.

5   Conclusion

This main paper is to address the problem of guaranteeing an acceptable level of QoS requirements for mobile users as they move from one location to another. This is achieved through reservation variables such as the current location and the moving direction that is presented with a set of attributes that describes the user mobility. In this scheme, mobiles are classified according to their reservation variables. Based on reservation variables a scheme that provides predictive QoS guarantees in mobile multimedia networks is proposed. The proposed scheme shows a great improvement of the resource utilization, the delay and the data loss. It is because our resource reservation scheme is more adaptive than existing resource reservation schemes. In our scheme, resources are classified as ones having priority to the new calls and ones having priority to the handoff calls based on reservation variable. We improve the dropping rate for the handoff connections by dynamically adjusting the amount of the reserved resources according to the amount of occupied resources. The determination of the optimal direction should be studied consecutively. Also further researches are required on their implementation and applications to the handoff and resource allocation strategies.

References

1. P. Agrawal, D. K. Anvekar, and B. Narendran, "Channel management policies for handovers in cellular networks," Bell Labs Tech. J., vol. 1, no. 2, pp. 97–10, 1996.
2. M. Naghshineh and M. Schwartz, "Distributed call admission control in mobile/wireless networks," IEEE J. Select. Areas Commun., vol. 14, pp. 711–717, May 1996.
3. O. T. W. Yu and V. C. M. Leung, "Adaptive resource allocation for prioritized call admission over an ATM-based wireless PCN," IEEE J. Select. Areas Commun., vol. 15, pp. 1208–1225, Sept. 1997.
4. C. H. Yoon and C. K. Un, "Performance of personal portable radio telephone systems with and without guard channels," IEEE J. Select. Areas Commun., vol. 11, pp. 911–917, Aug. 1993.
5. J. C. Lee and Y. S. Mun, "Mobile Location Estimation Scheme," SK telecommunications Review, Vol. 9, No. 6, pp. 968–983, Dec. 1999.
6. J. C. Lee, B. H. Ryu and J. H. Ahn, "Estimating Position of Mobiles by Multi-Criteria Decision Making," ETRI J., vol. 24, no. 4, pp. 323–327, Aug. 2002.

Mobile Station Positioning Using GSM Cellular Phone and Artificial Neural Networks

ZORAN SALCIC and EDWIN CHAN
Department of Electrical Engineering, Auckland University, 20 Symonds St., Auckland, New Zealand
E-mail: z.salcic@auckland.ac.nz

Abstract. In this paper, we describe a novel approach to mobile station positioning using a GSM mobile phone. The approach is based on the use of an inherent feature of the GSM cellular system (the mobile phone continuously measures radio signal strengths from a number of the nearest base stations (antennas)) and on the use of this information to estimate the phone's location. The current values of the signal strengths are processed by a trained artificial neural network executed at the computer attached to the mobile phone to estimate the position of the mobile station in real time. The neural network configuration is obtained by using a genetic algorithm that searches the space of specific neural network types and determines which one provides the best location estimation results. Two general methods are explored: the first is based on using a neural network for classification and the second uses function approximation. The experimental results are reported and discussed.

Keywords: cellular networks, positioning, artificial neural networks.

1. Introduction – Motivation for Research

The positioning systems using wireless communication can be traced back to the PULSE conference of 1968 [1], whose focus was on Automatic Vehicle Monitoring. Since then, many positioning systems have been proposed and used, and they will be described in Section 2. The advance of technology has changed the heavy, big systems into light-weight systems such as the Global Positioning System (GPS).

The development of positioning systems primarily focuses on the positioning of vehicles and, therefore, they are called Automatic Vehicle Monitoring (AVM) or Automatic Vehicle Location (AVL) systems. Although new systems such as the GPS system are used in positioning vehicles, they are not restricted to this application because of their small size, portability, and reasonable (low) cost.

Safety is the primary motivation for vehicle location. In the United States, the Federal Communications Commission (FCC) has adopted a Report and Order and Further Notice of Proposed Rule (NPRM) that creates rules to govern the availability of basic 911 services and the implementation of E911 (Enchanced-911) for wireless services. For basic 911 service, the Order requires all cellular, broadband PCS, and certain SMR licenses to transmit all 911 calls made from mobile handsets that have a code identification to a Public Safety Answering Point (PSAP) without any blocking or validation procedures. The Order also requires these carriers to provide certain E911 features which enable the PSAP to identify the location of the caller, including Automatic Number Identification (ANI) and Automatic Location Identification (ALI) within the required timetable [2]. Target accuracy is the ability to locate in latitude and longitude a wireless caller within 125 meters Root Mean Square (RMS).

While safety is the main motivation for wireless position location, other promising applications include navigation services, automated billing, fraud detection, roadside assistance, cargo tracking, and fleet management. Position location systems will provide new services and revenue sources for wireless carriers, greater crime-fighting capabilities for law enforcement personnel, and new methods for tracking people and parcels [3].

This paper presents an attempt at applying a new approach towards the positioning of a GSM-based mobile station. It uses radio signal strengths from serving and neighboring base stations that are continuously measured in each mobile station and applies them to a trained artificial neural network for positioning. As such, positioning is performed instantaneously (in real time). The paper is divided into the following sections. Section 2 presents a brief overview of the positioning approaches and explains the reasons for our approach based on the use of only signal strengths. Section 3 introduces the "Automatic GSM-based Positioning and Communication System" (AGPCS), which represents a context for the application of our positioning models. The AGPCS provides a framework to integrate positioning with communication in order to achieve an infrastructure for a number of exciting applications. A brief overview of artificial neural network features, which are used for positioning, and genetic algorithm features, which are used for the selection of the neural network models applied for positioning, is given in Section 4. Two categories of models are developed and presented in Sections 5 and 6. A classification model is used to try to determine the area in which the mobile station lies. The function approximation models are used to estimate mobile station position in two-dimensional space by estimating distances and/or angles of the mobile station from a number of base stations or to estimate directly two-dimensional coordinates of the mobile station. Experimental results of the application of developed models are presented in Section 7. The best positioning accuracy using the direct positioning method with ANN as the function approximator and adjusted ANN weights results in an average distance error just above 200 m. Section 8 presents conclusions and recommendations for future research directions.

2. A Survey of Positioning Systems

The positioning systems can be broadly classified into three major categories/classes:

1. systems using signal strength measurements,
2. systems using time of arrival of radio signal, and
3. systems using dead reckoning techniques.

The methods belonging to class 1 and 2 can be collectively called radio-location methods because they rely on the properties of radio signals.

An excellent overview of various positioning approaches and systems is given in [3]. That paper introduces major radio location-based positioning techniques and classifies them into categories according to their complexity: from basic techniques, such as triangulation, to the most advanced and complex techniques, such as those based on angle of arrival, time of arrival, and time difference of arrival of the radio signals.

The most accurate positioning today is achieved using the satellite-based global positioning system (GPS) [4]. The principle behind GPS, as a time-of-arrival system, is simple, although the implementation is quite complex. GPS uses precise timing within a group of satellites and transmits a spread-spectrum signal to earth on L-band. An accurate clock at the receiver measures the time between the signals leaving the satellites and arriving at the receiver. If at

least three satellites are visible to the receiver, triangulation can be used to find the receiver's location. Additional reference stations are used in GPS to improve accuracy further; this is called differential GPS. However, GPS has two important disadvantages. First, the information on position usually has to be transmitted to some other party requiring that the mobile station provides data communication facilities. Most often, it is provided using some sort of radio system-based data transfer, such as radio modems which communicate with a specialized radio system infrastructure (such as trunked radio), or using an existing public cellular system. In the former case, the problem of radio coverage is introduced, which requires investment in radio transmission systems. Further costs are incurred by data communication devices. In the latter case, the positioning and communication facilities are presently implemented by two separate devices. The second disadvantage of GPS is that it is usable only in the case of "clear sky", which makes it hardly usable in urban areas, mountainous terrain, and closed/covered space.

Dead reckoning methods locate a vehicle by computing its distance and direction of travel from a known fixed initial position. The distance measurements are made using a precision odometer and some compass type device to measure azimuth. Factors, such as side winds, change in tire pressure, and road conditions, can greatly affect the distance and direction measurement. Since the computed location depends upon all previous estimates, errors in location tend to accumulate and can quickly lead to sizable position error. To compensate for this, the system must be updated on a regular basis. This may be done manually, by proximity devices, or by comparing the vehicle's route with known or feasible routes. A system of this type, described in [5], uses such a technique corrected by central processor map correlation. This type of system has the advantage that it is not susceptible to problems associated with radio measurement techniques encountered in an urban environment.

Many methods and systems have been proposed based on the radio signal strength measurement [6–8] of a mobile station's transmitter by a set of base stations. Recently, adaptive schemes based on the use of cellular systems and on fuzzy logic [9], hidden Markov models, and pattern recognition methods [10, 11] have been used to estimate the position of mobiles. A study in [9] using computer simulation shows that the error between the exact position and the estimated position is in the range of 0 to 575 m. This is based on the assumption that 5 km separate the base stations from each other and the point-to-area terrain configuration has a 3 dB standard deviation from a normal signal strength distribution. The most recent simulation [12] is based on a multidimensional scaling technique that yields some very accurate results. A mobile's position is determined in a such way that the measured signal strength of a certain base station in the GSM system is best fitted to the known average signal strength at this point. The performance of the method was tested by simulation for different simulated scenarios [12], but no results from a real cellular environment have been reported.

Other positioning techniques have been proposed based on time angle of arrival (AOA) or time of arrival (TOA) of incident signals. In the AOA case, an estimate of position is made from base stations using a directional antenna to measure the AOA of incident signals [13] and requires complex adaptive high-resolution analysis techniques. If TOA is used [14], the three-dimensional position of the mobile station is uniquely determined by the intersection of three spheres. The major problem encountered with this technique is the requirement that all transmitters and receivers in the system have synchronized clocks. Otherwise, even a very small timing error could result in a very large position location error. Finally, the time difference of arrival (TDOA) method, which is based on the difference between the time at which the signal arrives at multiple base stations and the absolute time, is more practical for commercial

systems [15]. This technique requires only fixed base stations to have precisely synchronized clocks. Another limitation for all these techniques is that they rely on a direct line-of-sight path from the mobile station to the base stations. This is, however, not true for both urban and mountainous environments. The absolute time delay also undergoes multipath effects. The GSM-based equalization techniques rely on combining the resulting signal to produce the least error. This has the effect of distorting the time delay. However, what is required for positioning are multipath rejection algorithms. There have been a number of studies in this area, especially in the use of Least Mean Square techniques [16] and extended Kalman filters [17]. These techniques have proved to be of some value in removing multipath effects. Kalman filters in particular have also been used for the purpose of ?averaging? the mobile position from a set of fluctuating data, providing excellent results [18]. The investigation of the use of Kalman filters in this area, therefore, may lead to promising results.

Our paper presents the results of the research obtained so far as a part of the project that explores an Automatic GSM-based Positioning and Communication System (AGPCS) [19, 20] using a standard GSM mobile phone in a cellular environment. The fact that a standard GSM cellular phone is used for positioning imposes substantial constraints compared to the approaches that require additional precise equipment to support positioning. The main constraint is that only received signal strengths from the serving and a number of base stations are available for positioning purposes. First, we present the features of the cellular environment relevant to our approach to mobile station positioning. We also introduce the artificial neural network models that are used in our approach for positioning. Then, we describe our approach to positioning using two major general methods. The first approach is based on using a neural network for classification that determines the probable area the mobile phone is in. The second one uses the function approximation to determine either distances between the mobile phone and the neighboring base stations and triangulation to position the phone or the position (latitude and longitude, or X and Y coordinates) of the mobile station itself. Finally, we describe experiments and present results obtained from the application of our models on positioning in a real cellular environment and discuss the steps needed to improve the accuracy of our models.

3. AGPCS: GSM-Based Positioning and Communication System

GSM [21, 22] initially handled basic voice services and some emergency calling features, but has already added improvements to subscriber identity module (SIM) cards which contain a microchip with information on the caller. From the user point of view, the obvious difference between GSM and other cellular technologies is that GSM cellular phones operate only digitally, enabling both voice and data to be transferred directly digitally, without using modems, thus providing the backbone of the mobile communication network.

A variety of data services are offered in GSM. GSM users can send and receive data, at rates up to 9600 baud, to users on POTS, ISDN, Packet Switched PDN, and Circuit Switched PDN using a variety of access methods and protocols. Other data services include G3 facsimile and Short Message Services (SMS), which is a bi-directional service for short alphanumeric (up to 160 bytes) messages. Messages are transported in a store-and-forward fashion. For point-to-point SMS, a message can be sent to another subscriber and an acknowledgment of receipt is provided to the sender. SMS can be used in a cell-broadcast mode for sending messages such as updates of different sorts. Messages can also be stored in the SIM card for later retrieval.

The SMS service provides a basic means to transfer data used to estimate the position or coordinates of the mobile station.

Besides voice and data services, the GSM system provides data that might be used for radio signal strength measurements and positioning. The mobile station must continuously monitor the neighboring cells' perceived power levels. To do this, the mobile station receives a list of base stations (channels) on which to perform power measurements. The list is transmitted on the base channel, which is the first channel a mobile tunes in on when it is turned on. The GSM mobile station receives the downlink signal levels from the serving and up to six neighboring base stations in a discrete scale each 0.48 seconds. The GSM mobile station applies a complex signal-processing algorithm to determine the signal strengths. This information is part of the GSM system and is used in our system to estimate the position of the mobile station.

By using and integrating two inherent features of the GSM system (measurements of radio signal levels and ability to communicate directly digitally), we have proposed [19, 20] an Automatic GSM-based Positioning and Communication System (AGPCS). The AGPCS is a real-time system built on top of the GSM system and can be considered as an application layer to standard GSM. It performs the positioning of the mobile station in the coverage area of the GSM network. The AGPCS mobile station consists of the GSM mobile station (actually a handset) and a mobile computer connected to it. Depending on the power of the mobile computer, various degrees of intelligence and application complexity can be achieved within the AGPCS mobile station. The AGPCS mobile station performs continuous radio signal strength measurements and acquisition of measurements to estimate its position. The position is estimated using artificial neural network models that are based on current signal strength measurements, history of signal strength measurements, as well as some a priori knowledge of the environment, as will be shown in the following sections. The mobile computer collects signal strength measurements from serving and up to six neighboring base stations. Then, it either estimates the area in which the mobile station lies (using classification properties of neural networks), evaluates the distance between the mobile station and the neighboring base stations in order to use them in a triangulation model, or determines the position of the mobile station directly. This operation is performed in real time. This scenario leads to a self-positioning and communication system or SPCS. The SPCS is useful in applications in which the AGPCS mobile station and its user want to know their current position, and it can also transfer that information to the other parties.

In the second scenario, the mobile computer has a minimum of intelligence and input/output devices. It is used just to collect signal strength measurements, preprocess them, and transmit them to a network center (NC), where they are used to estimate position. A simplified version of the positioning model can run on the mobile computer and estimate the distances to the base stations, or position, which are then sent to the NC. The NC plays a supervisory role in the AGPCS system. Further refinement of position can be done and the corresponding database updated. The NC maintains data on the positions of a number of mobile stations and provides the means for presenting positions on a geographic map display. However, it can be used for various other purposes. The NC and a number of AGPCS mobile stations make up an AGPCS system. Obviously, the number of independent AGPCS systems or their architecture is not limited, because they depend only on the application requirements. Both scenarios involve the transfer of messages between the AGPCS stations or between the stations and the network center. This communication is performed without employing GSM voice channels. It is based on the short message service (SMS) that provides the exchange of short messages without using any additional interface equipment. Due to message latency, there can be delays in the

delivery of short messages. In that case, a direct link between the AGPCS mobile station and the network center can be established and information on signal measurements or position transferred immediately.

The first working versions of systems using AGPCS technology have been developed and tested. The AGPCS technology can be used for various applications, including control, as it may be easily incorporated into the standard hardware/software environments or used in the embedded form.

4. Neural Networks and Genetic Algorithms

It is known that only statistical models [23–28] can describe the received signal strengths at the mobile station. Resolving the variations of short-term fading, long-term fading, and the path-loss law is required to determine the position from the received signal strengths. The statistical models involve complex mathematical manipulations. Moreover, the statistical models are not universal. Extensive experiments are required to find appropriate expressions and parameters to describe the local behavior, which is time-consuming.

A much simpler, faster, and more flexible approach is to use artificial neural networks (ANNs) [29]. A wide range of neural network configurations and learning algorithms exists. In our work, we have placed a particular emphasis on the Multilayer Perceptron (MLP) neural network and its backpropagation learning algorithm. Obtaining an appropriate size is probably the most important task if a particular network topology is chosen. This will have profound effects on the performance of the network. Optimum network architecture (size) can be found using a trial-and-error approach or some optimization technique. The optimization technique used in our approach is a genetic algorithm [30].

Almost all of the approaches found in the literature to describe the behavior of the signal strength use a statistical model. They are usually founded on three assumptions: linearity, stationarity, and second-order statistics, with particular emphasis on Gaussianity. Yet, most, if not all, physical signals in real-life applications are generated by dynamic processes that are at the same time nonlinear, nonstationary, and nonGaussian. One way to analyze these processes is to use ANNs.

ANNs are a paradigm for the intelligent processing of information for some specific object-ive, e.g., classification, pattern recognition, decision-making, system behavior identification, and prediction. ANNs mimic human learning processes and therefore have great potential as adaptive learning systems [31]. ANN represents a method of synthesizing a mapping between input and output variables by learning a set of arc weights and node thresholds of a connec-tionist model based on training examples. They have been developed in a wide variety of configurations with some common characteristics, such as massive interconnection of simple computational elements. They are characterized by the model of their neurons, the connections between them, and the methods to train them to do a specific task.

ANNs have a number of important properties. Some of these properties are especially useful from the point of view of processes that take part in the estimation of the position of the mobile station from the signal strengths received from neighboring base stations [32]:

- ANNs are distributed nonlinear devices - they have the inherent ability to model under-lying nonlinearities contained in the physical mechanism responsible for generating the input data.

- ANNs have the potential to be fault-tolerant in the sense that the performance is degraded gracefully under adverse operating conditions.

- ANNs have the natural ability to adapt their free parameters to statistical changes in the environment in which they operate ANNs provide a nonparametric approach to nonlinear estimation. The term nonparametric is used in a statistical sense, meaning that no knowledge of the underlying probability distribution is required.

- ANNs in a supervised manner are universal approximators. Multilayer feed-forward networks are universal approximators in the sense that they can approximate any continuous input-output mapping to any desired degree of approximation given a sufficient number of hidden units.

For the purpose of mobile station positioning, we are interested in static networks whose output is a function of the current input. The multilayer perceptron (MLP) is the most widely used static neural network. The learning algorithm is the fast backpropagation, which is a supervised learning algorithm (backpropagation with a momentum) [29].

MLP is capable of approximating arbitrary nonlinear mappings, and given a set of examples, the backpropagation algorithm can be called upon to learn the mapping at the example point. However, there are a number of practical concerns. The first is the matter of choosing the network size. The second is the learning time. The final concern is the ability of the network to generalize: its ability to produce accurate results on new samples outside the training set [29]. Generalization is most heavily influenced by three parameters: the number of data samples, the complexity of the underlying problem, and the network size. The generalization property will also determine the learning time of the network.

A good generalization can only be achieved if optimum network architecture is found. In general, it is not known what size network works best for a given problem. Further, it is not likely that this issue will be resolved in the general case because each problem will demand different capabilities from the network. If the network is too small, it will not be capable of forming a good model of the problem. On the other hand, if the network is too large it will lead to overgeneralization and result in poor performance. There is a wide range of optimization algorithms to find the size of the network. All of them can be categorized into one of two general approaches:

1. A larger-than-necessary network is used initially and trained until an acceptable solution is found. After this, hidden units and weights are removed if they are no longer actively used. Methods using this approach are called pruning procedures.

2. A small network is used initially and then it grows additional units and weights until a satisfactory solution is found. Methods using this approach are called constructive procedures.

Another possible optimization technique that differs from pruning and constructive procedures used in our approach is the use of a genetic algorithm [30]. Genetic algorithms are a family of computational models inspired by evolution. These algorithms encode a potential solution to a specific problem on a simple chromosome-like data structure and apply recombination operators to these structures, in order to preserve critical information. In our approach, the input variables and the neural network structures, including the size of the network and the activation function used in the network, are encoded into the chromosome. The chromosomes are represented as binary strings. Each of these chromosomes is decoded into neural networks. The genetic algorithm process is based on a fundamental cyclic process, which consists of:

1. Creating an initial population of "genotypes" (genetic representation of the neural network).
2. Building neural networks ("phenotypes") based on the genotypes.
3. Training and testing the neural networks to determine how good they are.
4. Comparing the fitness of the networks and keeping the best ones.
5. Selecting those networks in the population that are better and discarding those that are not good enough.
6. Refilling the population back to the defined size.
7. Pairing up the genotypes of the neural networks.
8. "Mating" the genotypes by exchanging genes (features) of the networks.
9. "Mutating" the genotypes in some random fashion and returning to step 2.

The above cyclic process is called a generation and continues until some stopping criterium (required accuracy) is reached or the desired number of generations are performed. The best network obtained is the optimum network. The genetic algorithm is used to search for more relevant input variables from the provided input space, so the final network will have both optimal network architecture and a minimal number of input variables. All neural networks that have been investigated in our approach to positioning are 2-layer networks. The activation function in the hidden layer is the tangent hyperbolic, tanh, while the genetic algorithm determines the type of the activation function in the output layer. The fitness function used in step 4 is defined as the root mean squared error on the test set. Root mean squared error is calculated between the expected and actual neural outputs and is averaged across all output neurons, if more than one is employed.

5. Positioning Using Classification

Classification can be used to position the mobile station when the precise location of the mobile is not required. When using classification, the area of interest is divided into smaller sections and the identification of the section in which the mobile station can be found is the task of classification. The classification model and results of its application to the considered area are presented in this section.

5.1. CLASSIFICATION MODEL

ANN is used to perform the classification of the received signal strengths which represent input signals from a number of base stations' antennas. Experiments were performed in an area with a size of approximately 3 km by 4 km in which 4 base stations with ten antennas were located; three of them used directional antennas as described in the next section. The situation is illustrated in Figure 1. All coordinates are presented in NZMG (New Zealand Map Grid) format, which uses X and Y coordinates in meters and is directly related to the usual longitudes and latitudes.

The GSM mobile station provided the signal strength measurements from serving and up to six neighboring base stations' antennas. We limited this number to six signal strength values, which are collectively called a measurement record. These six values were divided into two groups of three measurements each with a time delay of around 2 seconds due to the limitation of equipment used. Despite this delay and time difference, we assumed that they were collected at the same time. By looking at the data collected at the training and
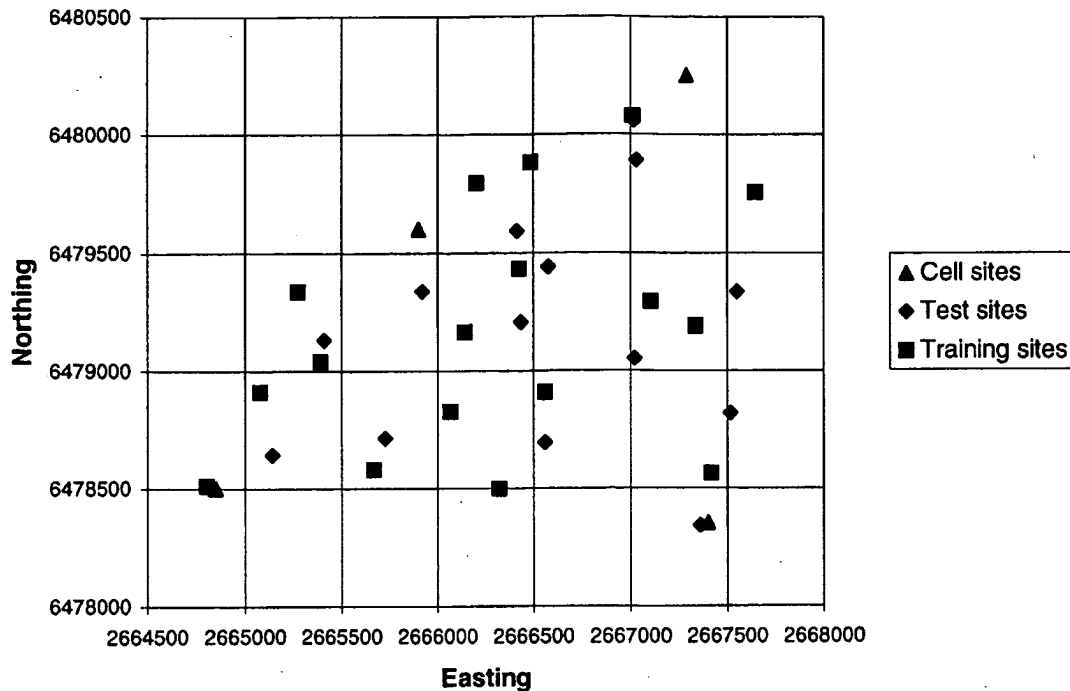
Figure 1. Illustration of the area used in the experiments.

testing sites, at most 4 of the 10 antennas appeared in the measurement record. The remaining readings were from other base stations outside the area included in the experiment. In the case of readings obtained from the base stations outside the considered area, we assumed that the signal strengths from remaining antennas not present in the record had value 0. This obviously represents a further approximation. Nevertheless, since the GSM mobile station does not provide the actual values of those measurements, we found it the only practical way to tackle the problem.

For classification purposes, the considered area was first divided into 3 smaller sections as illustrated in Figure 2. A neural network was developed for this classification problem and used to classify input measurement records. A second model was then developed by dividing the same area into 8 smaller sections as shown in Figure 3, and the performance of the two models was then compared. Each section was coded using the 1-out-C coding scheme in which C represents the number of categories present in the output.

ANN tries to learn from the input space (signal strengths of various antennas) and the corresponding output of the training data to predict the section the mobile station is in from the input space of the testing data. The predicted outputs in the test set are computed by passing the received signal strengths collected at the testing sites through the optimum network discovered by the genetic algorithm. The predicted outputs are then compared to the actual outputs. Accuracy of the classification for each site is the average correct classification made for all outputs for that site.

Two cases were considered:

1. The optimal network found by the genetic algorithm that contains a minimum number of inputs (from antennas that influence the classification process), and
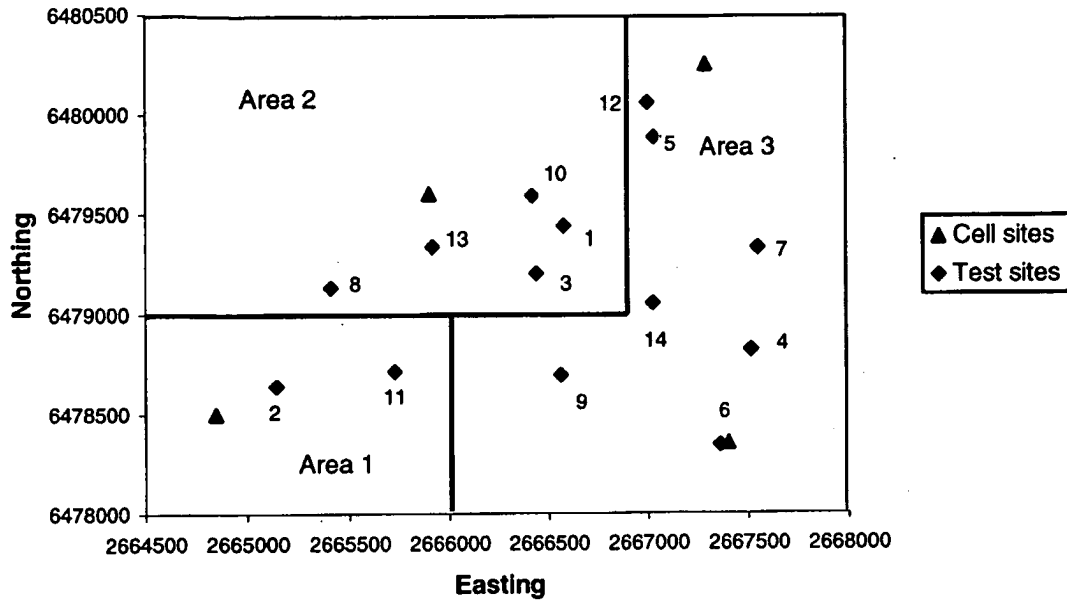
Figure 2. Division of the area into three sections.
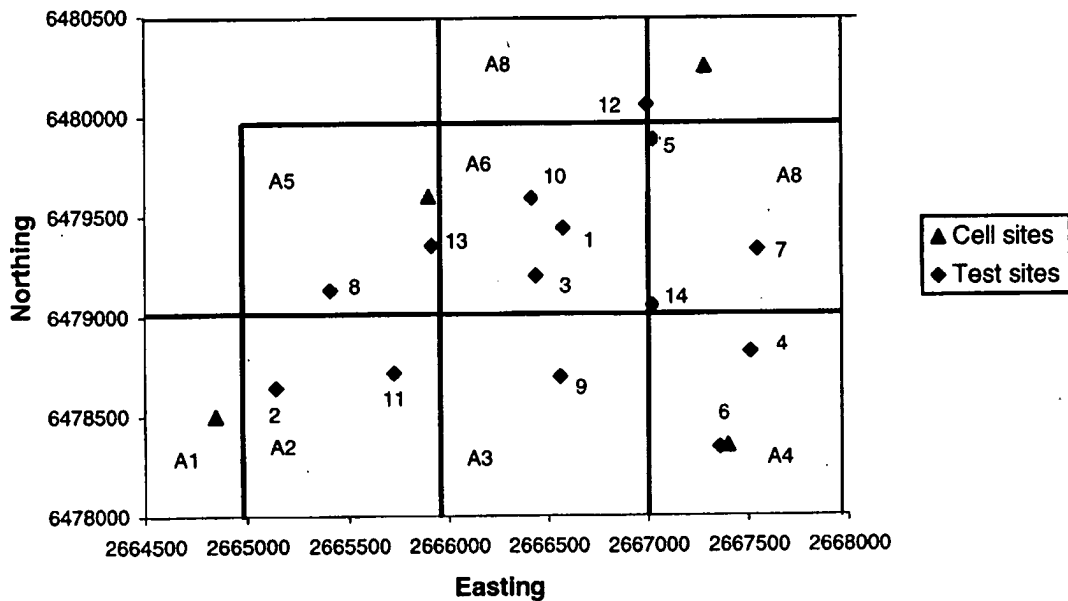
Figure 3. Division of the area into eight sections.

2. The optimal network found by the genetic algorithm that takes into account all inputs (all antennas in the considered area) regardless of their presence in the measurement records.

5.2. RESULTS OF CLASSIFICATION MODEL APPLICATION

In this section, we present the performance of classification models when the considered area was divided into 3 and 8 smaller sections. The sections were non-overlapping with sharp boundaries between them as shown in Figures 2 and 3. Our aim was to investigate how

Table 1. Results of the classification of the ANN with 3 and 8 outputs.

Site number	Area divided into 3 sections			Area divided into 8 sections		
	Desired	Correct class. rate (%, non-optimal no. of inputs)	Correct class rate (%, minimal no. of inputs)	Desired section	Correct class. rate (%, non-optimal no. of inputs)	Correct class. rate (%, minimal no. of inputs)
1	2	100	89.28	6	96.42	96.42
2	1	96.07	86.27	2	82.35	82.35
3	3	96.82	65.07	6	100	95.23
4	3	90.38	100	4	98.07	98.07
5	3	97.82	100	7	82.60	84.78
6	3	89.09	100	4	81.81	87.27
7	3	100	100	7	96.00	96.00
8	2	100	97.22	5	55.56	55.56
9	2	31.48	92.59	3	0	47.62
10	2	92.45	73.58	6	94.33	100
11	1	48.14	42.59	2	46.29	83.33
12	3	94.73	98.24	8	84.21	75.43
13	2	98.07	94.23	5	40.38	65.38
14	14	66.67	85.96	7	96.49	96.49

classification models position mobile stations in given areas and analyze the accuracy of these models. Network architectures obtained by the application of a genetic algorithm have a single hidden layer and the neuron hyperbolic tangent activation function in the hidden layer. The limitation imposed on the genetic algorithm was to investigate ANNs with up to 64 hidden neurons.

First, we considered the case in which the area was divided into 3 smaller sections. As a result of the genetic algorithm application, when received signal strengths from 10 antennas were used as input variables, the best network architecture contained 48 hidden neurons using tanh activation function and 3 output nodes with logistic activation function. The overall correct classification rate of the test set was 85.39%. If the optimized neural network was used, the genetic algorithm found that the model with only 5 input variables described the classification even better. The best network architecture had 5 inputs, 29 hidden nodes with tanh activation function, and 3 output nodes with linear activation function. The overall classification rate of the test set was 86.62%. The classification rates for all locations used in the experiment for both models are presented in Table 1.

The results of classification for positioning in Table 1 show that sites 9 and 11 were wrongly classified when the considered area was divided into 3 sections and all 10 input variables were used. The correct classification rate was fairly high for all locations other than these two sites, indicating that there was a high confidence in correct classification. When the ANN was used with an optimized number of input variables, the overall classification rate was improved and only site 11 had a classification rate that can be considered incorrect.

The second case is similar to the first one, but the considered area was divided into 8 smaller sections. The best neural network found for the model with 10 input variables was the network with 54 hidden nodes with tanh activation function and 8 output nodes with logistic activation function. The overall classification rate in this case dropped to 76.09%. If the optimized neural network was used, it again contained only 5 input variables, 49 hidden nodes with tanh activation function, and 8 output nodes with linear activation function. The overall correct classification rate of the test set was 80.43%. The results of classification for all 14 locations are shown in Table 1. When the area was divided into eight 1 km by 1 km sections, the overall classification rate in the test set was reduced to 76.09% for the ANN with 10 input variables and to 80.43% for the optimal ANN with 5 input variables. For a non-optimized network, classification for site 9 was completely incorrect and for sites 11 and 13 unsatisfactory, while for the optimized input variables case only site 9 was classified unsatisfactorily.

From the results of the application of ANN models, a number of observations can be pointed out. The correct classification rate in the test set improves when ANNs with an optimized number of input variables are used. Also, the performance of ANN with a lower number of outputs is better. Despite the sources of error (reduced accuracy), the results have proven the feasibility of the approach for mobile station positioning. This approach can be useful if the precise location of the mobile is not required, such as first estimate of the position or, for example, in an application such as the handover process. One of the possible improvements of the classification model is to introduce a degree of overlapping between sections to avoid misclassification. This approach will be analyzed in our future experiments.

6. Positioning Using Function Approximation

Another approach to achieve the mobile station positioning is to use ANNs as a function approximator. We analyzed two methods:

1. The received signal strengths are used to establish the relationship between the distance and angle of the mobile station and the antenna and indirectly take part in positioning. Two cases are considered further and used to position the mobile station:
   - estimation of distance and angle from the known antenna site.
   - triangulation, if at least three distances from three antennas belonging to three different base stations are known.

   Both these methods will be referred to as positioning with distance prediction.
2. The received signal strength input variables are used as in the preceding classification problem to model directly the two-dimensional coordinates of the mobile station. This method will be referred to as direct positioning.

6.1. POSITIONING USING DISTANCE PREDICTION

Using this approach, we first established the relationship between the received signal strengths and the distances to the experimental sites for each antenna. The best network architecture discovered by the genetic algorithm for each antenna was first found. Then, the predicted distance, the actual distance, and the distance error were determined for each test site. The final distance for a particular site was found by running the input data for that site through the neural network for that antenna and averaging for the number of samples taken.

Table 2. Average absolute angle error for all base stations.

Base station	A	B	C	D
Average absolute angle error (°)	6.15	14.80	6.30	17.74

Table 3. Dependence of average angle error on the distance.

Base station	A	B	C	D
Angle (°)	0.04	92.94	6.10	6.41
Distance (m)	1912	42.78	2515	1928

Another analysis we performed was to explore the existence of a relationship between inputs into the neural network and the angles between the mobile station and corresponding base stations. An analysis was performed based on the predicted coordinates from the model with the optimized number of input variables together with the correctional network. The actual distance of the mobile to each antenna and the error of the predicted angle were examined to see if a relationship existed. The predicted angles can be used later both as an input to other neural networks and to position the mobile directly. The predicted angle of the mobile station to the main beam of each antenna was used together with the signal strength to model the distance between the mobile and each antenna. The reason for this is because directional antennas are installed in the network and the radiation pattern is not identical in all directions. A two-dimensional problem is assumed.

The best direct positioning neural network with optimized inputs and correctional network was used to calculate predicted angles of the estimated positions to the base stations. Apart from some outliers, the predicted angles were quite accurate. The average absolute angle error ranged from 6.15° to 17.74° as shown in Table 2.

If we look at the dependence of the average angle error on the distance to the base station, it decreases as the distance increases. As an example, the estimated position for a typical site 6 as a function of the distance from the base stations is illustrated in Table 3.

Since the predicted angles showed a relatively high degree of accuracy, they were used as inputs in another neural network to position the mobile station. The predicted angles could also be used for the estimation of the initial position of the mobile station, as a part of the other positioning algorithm.

The relationship between the received signal strength and the distance from the mobile station to the base station was modeled for each antenna. Using a genetic algorithm, different neural networks were obtained as optimal for each antenna. Predicted angles for each test site were passed through the developed networks with the received signal strengths to output the predicted distance of the mobile station from a particular antenna. Table 4 illustrates the results obtained for site 1 and Table 5 gives the average absolute errors for all sites.

We applied and analyzed two approaches to position the mobile station using distance prediction:

1. Using the estimated distance and estimated angle of the mobile station to a particular antenna.

Table 4. Results of neural network application for site 1.[a]

Antenna	Predicted distance (m)	Actual distance (m)	Absolute error (m)
A3	1496	1077	419
B1	1227	1365	138
B3	1297	1365	68
C1	1217	1969	752
C2	1769	1969	200
C3	1393	1969	576
D	1033	697	305
Average distance error			351

[a] Antennas with the same initial letter are at the same site (base station).

Table 5. Average absolute errors for distance prediction for all sites.

Site	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Error	351	767	173.5	498.8	299.5	670.8	483.3	483.9	305.2	294.7	304.3	653.3	388.1	463.5

2. Using triangulation where distances to 3 antennas from three different base stations are required to position the mobile.

Results of both these approaches are presented below.

Positioning using estimated distance and estimated angle

The distance between the mobile station and a particular antenna and the angle of the mobile station to that antenna can position the mobile station. Since more than one antenna were present at each base station, the final calculated position was obtained by averaging the positions from all estimated positions. The final estimated positions for all sites are shown in Table 6.

Positioning using triangulation

The predicted distances from the mobile station to the three base stations were used to estimate the position of the mobile station using triangulation. Table 6 shows errors in the estimated positions (in meters). Obviously, this model produces excessive errors.

6.2. DIRECT POSITIONING

This model cascades two neural networks in a feed-forward manner as shown in Figure 4. The input space in the first network was the signal strength of the various antennas, the same as in the classification approach. Zeros were used to represent situations in which no antenna was found in the measurement record (not one of the neighboring antennas). The output of this network represents the position of the mobile station in the NZMG coordinates (Northing, Y, and Easting, X). The training and validation procedures for this ANN were the same as for the classification model, but the outputs were the coordinates of the mobile station. After the

Table 6. Position estimation using estimated distances and angles and triangulation.

Location	Model using distances and angles	Triangulation model
	Distance error (m)	Distance error (m)
1	447.66	700.1
2	168.90	1827
3	95.68	235.7
4	298.20	419.9
5	348.00	244.2
6	642.86	812.6
7	461.43	606.4
8	292.84	819.7
9	537.96	382.8
10	281.86	895.6
11	290.00	599.0
12	604.65	2062
13	393.00	788.3
14	378.00	291.8
Average distance error (m)	374.36	763.2
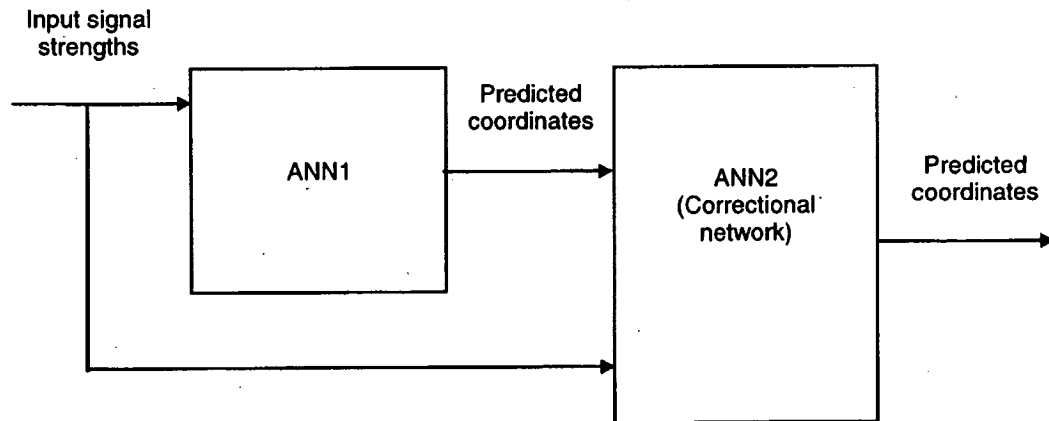
Figure 4. Cascading two neural networks to achieve direct positioning.

training session was completed, the predicted NZMG coordinates of this ANN in the training and testing sets were used together with the original signal strengths to train and validate a second ANN. The outputs of the second network were again the NZMG coordinates that represent the mobile station position. This second network can be considered as a correctional network.

Table 7. Experimental results of the direct positioning method.

Model	Average absolute error (m)	Minimum error (m)	Maximum error (m)	Main network[a]	Correctional network
Non-optimal without correctional network	313.95	127.34	659.49	H.N. = 45 tanh O.N. = 2 linear Inputs = 10	
Non-optimal with correctional network	290.38	111.02	505.72		H.N. = 44 tanh O.N. = 2 linear Inputs = 10
Optimal without correctional network	301.18	142.74	650.60	H.N. = 51 tanh O.N. = 2 logistic Inputs = 7	
Optimal with correctional network	278.79	74.33	528.25		H.N. = 50 tanh O.N. = 2 logistic Inputs = 7

[a] H.N. = hidden node; O.N. = output node.

Both ANNs were found by using a genetic algorithm. Two situations, as in the classification method, were considered:

1. An optimal neural network that considered all input variables and
2. An optimal neural network with the optimized (minimized) number of inputs. The final predicted outputs (coordinates) for each site were found by simple averaging over the number of samples taken at that location.

The results of applying all direct positioning models are summarized in Table 7 for comparison purposes. These results were obtained for 14 test sites.

The cascade of two neural networks tried to achieve a feedback architecture to analyze whether performance will improve. Although all networks were individually trained, they operate as a single network. From the experiments, it is obvious that the networks with the optimized number of inputs improve positioning accuracy. Also, the correctional networks improve the average absolute error slightly, but reduce the maximum and minimum errors significantly.

Recently, we noticed that besides the network architecture, presented by the number of input variables, hidden layers, and hidden neurons, and the type of activation functions used, some neural network parameters also had a large impact on model accuracy. These parameters included weight initialization range, learning rate, and momentum. First, a better weight initialization range was found. We made an attempt to adjust neural network parameters and further improve the results of the best-found network. Then, the network model based on that configuration was trained. The initial value of the weight initialization range was [−0.3, +0.3]. The motivation for starting with small weights was that large weights tend to prematurely saturate units in a network and render them insensitive to the learning process. However, if the

Table 8. Comparison of all analyzed models in terms of average distance error.

Model	Average distance error (m)
Direct positioning – Optimal with correctional network and adjusted weight initialization	205.32
Direct positioning – Optimal with correctional network	278.79
Direct positioning – Non-optimal with correctional network	290.38
Direct positioning – Optimal without correctional network	301.18
Direct positioning – Non-optimal without correctional network	313.95
Predicted distances and angles	374.46
Triangulation	763.20

weights are too small, a similar problem can occur. Therefore, it is hard to determine what the initial weight value should be. A strategy for choosing the value of initial weight for avoiding premature saturation was suggested in [33] and was applied to our best-found network with the correctional network. The average distance error was improved further for more than 25% to 205.3 m, indicating that the better ANN training techniques and tuning of the parameters should be investigated.

7. Analysis of Results and Discussion

The total number of antennas was 10. The GSM service provider (BellSouth N.Z.) furnished the data on the antennas. The area used in experiments was, according to the definition of [28], predominantly a suburban one with a mixture of light urban areas. The light urban category corresponds to the fringes of the Central Business District and to suburban shopping centers, where buildings are usually never higher than two or three levels. Total number of measurements taken for each training and test site was around 50.

Table 8 is used to compare average absolute errors of the estimated positions for all applied models. As can be seen, the best results were achieved using the direct positioning method with the optimal number of inputs and correctional neural network. Generally, the methods with direct positioning are less demanding computationally.

All results shown in this paper were obtained as average values by passing a number of measurements through corresponding neural networks. The actual positions of all experimental sites were obtained using a GPS unit. Therefore, the error of the GPS system was incorporated into the current models. Its influence on our final results could not be analyzed. Moreover, our analysis assumed the positioning problem to be two-dimensional, introducing another source of error.

Despite the number of potential sources of error, the current investigation has pioneered a new, fast, and flexible paradigm for positioning using neural networks avoiding complex statistical manipulation of the signal variations.

The results reported in this paper were based on 17 training and 14 testing sites in a 3 km × 4 km area. Furthermore, the project team knew of only 10 antennas on 4 base stations at the time of the experiments. Actual signal strength measurements were limited to the capabilities of a standard GSM mobile station. Current input data had no preprocessing. As our primary goal was positioning, all data samples were collected in a slowly moving environment. Thus, the models cannot be verified for faster moving mobile stations.

A genetic algorithm search to find the best candidate ANNs usually runs 1–2 hours on a Pentium II-300 PC when restricting the type of ANN to MLP BP and the number of hidden neurons to 64. However, once its architecture is determined, the training of the ANN takes no longer than two minutes. Estimation of the position for a new measurement record is practically instantaneous.

Future research directions have been determined and are currently being studied. They include a much better data acquisition system and a more accurate and automated system for the input of actual positions and measurements, which is necessary to both train and validate models. Also, other neural network models and training mechanisms will be further investigated, especially those that provide better tuning of the networks and shorter training times. Another area in which the models described in this paper can be used is to model specific areas or streets where handover problems are constantly encountered. If the cellular network detects that the mobile station is constantly handed-over between 2 or more base stations, it can use the neural network models to detect the position of the mobile to reduce the load of the network.

8. Conclusion

This paper presents the results related to the positioning of the GSM mobile station using only the information present in the mobile station. As such, this positioning system can be used for both the positioning of the vehicles or mobile stations and the communication with the other parties in the system. Positioning of the mobile station was achieved by using signal strengths measured at the mobile station and then processed by artificial neural networks either in the mobile station itself or at the remote network center. Two major approaches for positioning were presented: using a neural network as a classifier that attempts to determine in which area the mobile station lies and using a neural network as a function approximator to determine the position of the mobile station. Two further refinements of function approximation were analyzed and compared: 1. direct positioning that estimates the position in terms of two-dimensional coordinates and 2. indirect positioning that calculates the estimated position from the estimated distances of the mobile station from the base stations (or antennas) and/or angles to the base stations. Our current experiments show that the model which directly estimates the position of the mobile station using a minimal number of inputs relevant for positioning and the correctional neural network performs the best. Some of the directions for future research were also outlined in this paper. Results obtained so far are encouraging and serve as a good basis for future research.

References

1. S.H. Roth, "History of Automatic Vehicle Monitoring", *IEEE Transactions on Vehicular Technology*, Vol. VT-26, No. 1, pp. 2–6, 1977.
2. IEEE Vehicular Society News, Vol. 44, No. 1, p. 6, 1997.
3. T.S. Rappaport, J.H. Reed and B.D. Woerner, "Position Location Using Wireless Communication on Highways of the Future", *IEEE Communication Magazine*, Vol. 34, No. 10, p. 33, 1996.
4. E. Kapplan, *Understanding GPS: Principles and Applications*, Artech House, 1996.
5. T.W. Lezniak, R.W. Lewis and R.A. McMillen, "A Dead Reckoning/Map Correlation System for Automatic Vehicle Tracking", *IEEE Transactions on Vehicular Technology*, Vol. VT-26, No. 1, pp. 47–52, 1977.
6. Figel et al., "Vehicle Location by a Signal Attenuation Method", *IEEE Transactions on Vehicular Technology*, Vol. VT-18, pp. 104–109, 1969.
7. G.D. Ott, "Vehicle Location in Cellular Mobile Radio Systems", *IEEE Transactions on Vehicular Technology*, Vol. VT-26, pp. 43–46, 1977.
8. M. Hatta and T. Nagatsu, "Mobile Location Using Signal Strength Measurements in a Cellular System", *IEEE Transactions on Vehicular Technology*, Vol. VT-29, pp. 245–252, 1980.
9. H.L. Song, "Automatic Vehicle Location in Cellular Communication Systems", *IEEE Transactions on Vehicular Technology*, Vol. 43, pp. 902–908, 1994.
10. O. Kennemann, "Pattern Recognition by Hidden Markov Models for Supporting Handover Decisions in the GSM System", in *Proc. 6th Nordic Seminar Dig. Mobile Radio Comm.*, Stockholm, Sweden, 1994, pp. 195–202.
11. O. Kennemann, "Continuous Location of Moving GSM Mobile Stations by Pattern Recognition Techniques", in *Proc. 5th Int. Symp. Personal, Indoor, Mobile, Radio Comm.*, Den Haag, Holland, 1994, pp. 630–634.
12. M. Hellebrandt, R. Mathar and M. Scheibenbogen, "Estimating Position and Velocity of Mobiles in a Cellular Radio Network", *IEEE Transaction on Vehicular Technology*, Vol. VT-46, No. 1, pp. 65–71, 1997.
13. S.V. Shell and Gardner, "High Resolution Direction Finding", *Handbook of Statistics*, Vol. 10, Elsevier, pp. 755–817, 1993.
14. B.T. Fang, "Simple Solutions for Hyperbolic and Related Position Fixes", *IEEE Trans. Aerospace and Elec. Systems*, Vol. 26, No. 5, pp. 748–753, 1990.
15. Y.T. Chan and K.C. Ho, "A Simple and Efficient Estimator for Hyperbolic Location", *IEEE Trans. Signal Processing*, Vol. 42, No. 8, pp. 1905–1915, 1994.
16. H. So and P. Cheng, "Target Localization in the Presence of Multipath", *Electronic Letters*, Vol. 29, pp. 293–294, 1993.
17. R. Iltis, "Joint Estimation of PN Code Delay and Multipath Using the Extended Kahlman Filter", *IEEE Transactions on Communications*, Vol. 38, pp. 1677–1685, 1990.
18. M. Pent, M.A. Spirito and E. Turco, "Method for Positioning GSM Mobile Stations Using Absolute Time Delay Measurements", *Electronics Letters*, Vol. 33, No. 24, 2019–2020, 1997.
19. Z. Salcic, "AGPCS – An Automatic GSM-based Positioning and Communication System", in *Proceedings of the 2nd International Conference on Geocomputation*, Dunedin, New Zealand, August 1997, pp. 147–154.
20. Z. Salcic, E. Chan, H. Fok and A. Ip, "Integrating Positioning, Communication and Control in A Mobile Computing Environment", in *Proc. of the Fifth ANZ International Conference on Intelligent Information Processing Systems ICONIP '97, Dunedin*, Springer Verlag, 1997, pp. 931–934.
21. J. Scourias, "Overview of the Global System for Mobile Communications", University of Waterloo, Canada, Technical Report, May 1995.
22. S.H. Redl, M.K. Weber and M.W. Oliphant, *An Introduction to GSM*, Artech House, 1995.
23. J.D. Parsons, *Mobile Communication Systems*, Blackie: Glasgow; Halsted Press: New York, 1989.
24. R. Steele, *Mobile Radio Communications*, Pentech Press: London, 1992.
25. M. Hata, "Empirical Formula for Propagation Loss in Land Mobile Radio Services", *IEEE Transaction on Vehicular Technology*, Vol. VT-29, No. 3, pp. 317–325, 1980.
26. M. Ibrahim and J.D. Parsons, "Signal Strength Prediction in Built-Up Areas", *IEE Proceedings*, Vol. 130, No. 5, pp. 377–384, 1983.
27. R.H. Clarke, "A Statistical Theory of Mobile-Radio Reception", *The Bell System Technical Journal*, pp. 957–1000, 1968.

28. G.B. Rowe, "A Land Mobile Radio Coverage Area Prediction Model for New Zealand", The University of Auckland, Ph.D. Thesis, 1984.

29. P.J. Braspenning, F. Thujisman and A.J.M.M. Weijters, *Artificial Neural Networks: An Introduction to ANN Theory and Practice*, Springer Verlag, 1995.

30. D.E. Goldberg, *Genetic Algorithms: in Search, Optimization and Machine Learning*, Addison-Wesley, 1989.

31. M.A. El-Sharkawi, "Neural Network's Power", *IEEE Potentials*, pp. 12–15, 1996.

32. D.R. Hush and W.G. Horne, "Progress in Supervised Neural Networks – What's New Since Lippmann?", *IEEE Signal Processing Magazine*, 1993.

33. L.F.A. Wessels and E. Barnard, "Avoiding False Local Minima by Proper Initialization of Connections", *IEEE Transaction on Neural Networks*, Vol. 3, pp. 899–905, 1992.

Zoran Salcic received his B.E., M.E., and Ph.D. degrees in Electrical Engineering from Sarajevo University in 1972, 1974, and 1976, respectively. Part of his postgraduate work was conducted at City College of City University of New York in 1974 and 1975. He worked as an assistant professor and an associate professor at the Sarajevo University and Czech Technical University, Prague. Between 1985 and 1990, he was Deputy CEO and then CEO of the Institute for Computer and Information Systems of Energoinvest Corporation, Sarajevo. He has been with Auckland University since 1994. He has published over 80 technical papers, numerous technical reports, and six books. Furthermore, he has led a number of research, development, and large engineering projects. His current research interests include custom-computing machines, field-programmable logic and its applications in embedded and recon-figurable systems, complex digital systems design, automatic vehicle location, and related applications of mobile computing.

No photo available

Edwin Chan received his B.E. and M.E. degrees in Electrical Engineering from the University of Auckland in 1996 and 1998, respectively. His research interests are wireless communications and mobile computing. Since 1998, he has been with Hong Kong Telecom.